Final Technical Summary



# Parametric Study of Diffusion-Enhancement Networks for Spatiotemporal Grouping in Real-Time Artificial Vision

R.K. Cunningham
A.M. Waxman

6 April 1993

**Lincoln Laboratory**
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
*LEXINGTON, MASSACHUSETTS*

Prepared for the Department of the Air Force under Contract F19628-90-C-0002.

This report is based on studies performed at Lincoln Laboratory, a center for research operated by Massachusetts Institute of Technology. The work was sponsored by the Air Force Office of Scientific Research under Contract F19628-90-C-0002.



The ESC Public Affairs Office has reviewed this report, and it is releasable to the National Technical Information Service, where it will be available to the general public, including foreign nationals.

This technical report has been reviewed and is approved for publication.

FOR THE COMMANDER

Gary Titungian
Administrative Contracting Officer
Directorate of Contracted Support Management

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
LINCOLN LABORATORY

# PARAMETRIC STUDY OF DIFFUSION-ENHANCEMENT NETWORKS FOR SPATIOTEMPORAL GROUPING IN REAL-TIME ARTIFICIAL VISION

FINAL TECHNICAL SUMMARY
TO THE
AIR FORCE OFFICE OF SCIENTIFIC RESEARCH

*R.K. CUNNINGHAM*
*A.M. WAXMAN*
*Group 21*

**6 APRIL 1993**





LEXINGTON MASSACHUSETTS

## ABSTRACT

Spatiotemporal grouping phenomena are examined in the context of static and time-varying imagery. Dynamics that exhibit static feature grouping on multiple scales as a function of time and long-range apparent motion between time-varying inputs are developed for a biologically plausible diffusion-enhancement bilayer network. The architecture consists of a diffusion layer and a contrast-enhancement layer coupled by feedforward and feedback connections; time-varying input is provided by a separate feature extracting layer. The model is cast as an analog circuit that is realizable in VLSI, the parameters of which are selected to satisfy a psychophysical database of the following long-range apparent motion phenomena: gamma motion of a single light, smooth motion between two lights, reverse motion, split and merge among three lights, Ternus motion among multiple lights, and peripheral motion. The relation between motion on a uniform network (i.e., cortex) and inputs to a nonuniform sampling array (i.e., retina) are discussed in the context of a logarithmic scaling of space. A new interpretation of short- and long-range visual motion systems is introduced.

# PREFACE

This is the Final Technical Summary of the MIT Lincoln Laboratory parametric study of diffusion-enhancement networks for spatiotemporal grouping in real-time artificial vision. This report contains the selected results of many simulations performed during the past year, utilizing the network architecture and simulation tools developed during the first two contract years [1,2]. Discussed in this report are simulations of new variants of long-range apparent motion experiments. Also included are sections relating this work to other work, sections covering the network's inputs and outputs, and discussions of the role of the many forms of visual grouping phenomena and motion detection.

## ACKNOWLEDGMENTS

Robert Cunningham would like to acknowledge many conversations with Mike Seibert, Ennio Mingolla, and David Fay, which pointed the authors to valuable references and offered insights that significantly improved this investigation.

# TABLE OF CONTENTS

## TABLE OF CONTENTS
(Continued)

# LIST OF ILLUSTRATIONS

# 1. INTRODUCTION AND HISTORY

The phenomenon of long-range apparent motion between two lights separated in space and time was first demonstrated in 1875 by Exner [3] and is the simplest example of a dynamic spatiotemporal grouping process. Many phenomenological theories have been proposed, but few neurodynamic formulations exist. Because such grouping is preattentive, requiring no knowledge, it must ultimately be achieved in a simple neural architecture.

Proposed in this report is a diffusion-enhancement bilayer (DEB) that supports the spatiotemporal grouping process and can explain many effects reported in visual [4,5], auditory [6,7], and tactile [8] senses. The predecessor of the proposed model obtained fixation points for object recognition in sequences of static imagery [9]. This early model, known as the neural analog diffusion-enhancement layer (NADEL), was quickly recognized as relevant to static grouping and long-range apparent motion [10–12]. The NADEL and the DEB share a similar structure: a diffusion process coupled with a local maxima detector that directs feedback to the diffusion layer. The diffusion process acts as a long-range signaling function, which is accomplished in vivo by long neural axons, laterally connected neurons, diffusion or propagation of ions, or by other methods. The maxima detector suppresses the less salient input features while enhancing the more salient; it is accomplished by a recurrent neural network [13]. Feedback enables long-range interaction in limited precision neurons, but the two models differ slightly: DEB is distributed positive feedback, while NADEL is isolated or center-surround feedback. Finally, DEB input rises and falls, while NADEL input is constant. The result of these modifications is that while the NADEL model replicates the path of apparent motion, but not the direction, the DEB model replicates both and can explain a wider variety of stimuli than the NADEL.

To further explain the DEB, the reader should understand the context of this work and some of the biological and psychophysical results that led to the development of this model. Thus a literature review is presented before the model itself is described. After describing the model and a possible relationship to physiology, numerical simulations of a large variety of long-range motion phenomena are performed in which stimuli are presented in the presence of "endogenous" noise. In these simulations the network produces the perceived response while suppressing the endogenous noise.

## 2. VISUAL MOTION SYSTEMS: THEORIES AND MODELS

### 2.1 Motion System Divisions

In 1974 Braddick [14] proposed a delineation for two motion systems that he and Anstis later more fully defined [15,16]. Several properties differentiated the two systems. System 1, historically named short-range, was thought to detect movement over small distances (< 15-min visual arc) and short times (80- to 100-ms interstimulus intervals—ISIs—or the time between two successive stimuli) [14], produce motion aftereffects, not respond to color but to like contrast [17,18], and not respond to dichoptic presentation (presenting alternating stimuli to alternating eyes) [7].[1] System 1 is associated with random-dot kinematogram stimuli. System 2, historically named long-range, was considered to operate over broad regions of space (several degrees of visual arc) and time (up to 500 ms) [35], respond well to dichoptic presentation and color, be contrast insensitive, and not exhibit motion aftereffects. System 2's associated stimuli are small numbers of dots or sparks or simple shapes.

The two systems are not as separable as was once thought. System 1's spatial limit can be increased if the spatial frequency of the stimuli are decreased [19], while System 2 could always operate over a short spatial scale as well. In contrast with the original delineation, kinematograms (System 1 stimulus) in which black/white dots are replaced with isoluminant red/green dots have recently been shown to produce apparent motion [20], provided the ISI is not filled with a dark frame as in the original study [18]. In addition, a weak form of motion aftereffects has now been discovered in System 2 [21]. Braddick's observation, however, that System 1 does not produce apparent motion when successive frames are dichoptically presented still appears to be true [14,22], and contrast reversal still seems to stimulate only System 2 [23].

More recently, Cavanagh and Mather proposed a model that is "a concatenation of a common mode of initial motion extraction followed by a general inference process" [22]. This model introduces two new classes: a first-order motion system, which is defined by its ability to respond "to the displacement of first-order differences in luminance and perhaps colour" and a second-order motion system, which responds "to displacement of second-order differences in luminance or colour [such as texture, local motion, or disparity], even in the absence of first-order differences." The second-order system elaborates on the ability to perceive motion in the drift-balanced random stimuli of Chubb and Sperling [24]. Although Cavanagh and Mather reject the system divisions of Braddick and Anstis, their second-order system can be seen as incorporating emergent contours (e.g., the

[1] When stimuli presented to one eye are able to interact with stimuli presented to the other eye, the interaction must take place after retinal signals come together in visual cortex. Short-range motion requires presentation of alternating stimuli to the same eye, implicating motion analysis that takes place before retinal signals converge.

boundary contours of Grossberg and Mingolla's Boundary Contour System [25,26]) as input to a short-range motion system.

The DEB model reported here suggests an alternative interpretation of the classical short-range/long-range dichotomy that integrates some of the ideas of Cavanagh and Mather [27,28,1]. This interpretation contrasts a local motion process (perhaps of luminance as in the original Braddick/Anstis division or luminance and second-order stimuli as in Cavanagh/Mather) with a spatiotemporal grouping process that generates moving activity patterns between widely spaced stimuli. The input stimuli can be either first- or second-order, but it is the moving activity patterns that feed a later local motion system. In this interpretation, the local motion system samples continuous movement (thereby operating effectively on stroboscopic discrete movements) either provided by smoothly moving stimuli (short-range system) or smoothly moving activity patterns of the spatiotemporal grouping process (long-range system). In this report the spatiotemporal grouping process is explicitly modeled via coupled diffusion and enhancement networks with the result that many effects attributed to the long-range system can be explained in terms of this spatiotemporal grouping process.

All the motion system divisions attempt to account for psychophysical data, suggesting that motion can be experienced both locally and globally in stimuli that differ primarily in luminance levels and globally in more complex stimuli. They also attempt to account for physiological data suggesting that motion processing occurs at multiple visual levels, in multiple cortical paths. Regardless of how psychophysical percepts map onto the physiology, there must be an evolutionary advantage to these levels and paths, and it is likely that they perform several different types of motion processing. Furthermore, the paths converge and diverge several times, so it should be expected that visual motion systems are not perfectly distinguishable in perceptual tests.

### 2.2 Short-Range Models

The most common motion models address the short-range process, sampling continuous motion. There appear to be two variants—luminance-based models, which are associated with a first-order short-range process, and feature-based models, which are associated with first- and second-order short-range processes. Luminance-based models compare local luminance at one location in time to local luminance at another location at some later time. These models fall into two classes: correlation models [29–33] and gradient models [34–36]. Many of the correlation models are used to describe the behavior of lower-level animals (flies, rabbits), but some also describe human perception. These models compute the velocity of image intensity at a rate tuned by the temporal and spatial properties of the correlation, and multiple scales provide a range of speeds.

Gradient models examine differential changes in intensity, or some filtered version of it, over space and time. Both correlation and gradient models typically assume that brightness patterns are convected along with image motion, which is often not the case (e.g., objects that traverse a luminance gradient cannot be accurately tracked by these models but they can by human observers). One suggested implementation [35] computes "edge" directions to within 180 deg and speed at the

rate of the tuned spatial/temporal receptors. By combining the output of this computation with several directions and several scales, directional and speed ambiguities are removed.

Feature-based models track features such as edges or corners, texture discontinuities, illusory contours, or motion discontinuities by postulating the temporal growth and decay of a Gaussian activity wave in response to a transient input [37,38,12]. For noninteracting inputs (i.e., a Gaussian wave with scale smaller than the spacing of inputs), this process provides a means to extract directly the speed of moving features. When the features are close to one another the waves do interact, and in so doing they interpolate the trajectory between inputs. These models directly compute local velocities.

Both luminance- and feature-based models are intended to sample continuous motion, but their spatial and temporal sampling will permit small, discontinuous motion to be interpreted as continuous motion.

### 2.3 Long-Range Models

Cavanagh and Mather [22] have suggested that large-scale versions of short-range motion models may suffice as a model of long-range apparent motion. Unfortunately, such models cannot account for the strong percept of object localization in phi motion, the long-range interactions of small objects (with only high spatial frequencies), or the potentially nonlinear velocities perceived in apparent motion experiments.

Several models specific to long-range motion have been developed. Ullman presents a 2-D visual-apparent motion model [39], where objects are assigned relative affinities such as relative object length, interobject distance, and relative orientation, which relate the experimentally determined probability that two objects will group. After assigning affinities to all possible local pairs, the affinities are summed and minimized. The model performs well because it is based on empirically determined probabilities. It is unclear that this model represents any underlying biological process and is better appreciated as a statistical summary of experimental data.

The DEB is a 1-D model that treats incoming features as sources to a spatiotemporal grouping process. The result of this grouping process is an activation profile, the maxima of which induce a localized percept that can be tracked by one of the abovementioned short-range models.

Another 1-D model of long-range apparent motion was recently proposed by Grossberg and Rudd [40]. The basic model elements that are responsible for creating continuous motion paths from spatially disparate inputs are related to those used in the DEB model and its precursor, the *NADEL*. Essentially, localized inputs (e.g., flashes of light) are assumed to excite a spatially extended Gaussian activation pattern of fixed scale. By combining a preprocessing stage, which detects spatial gradients of brightness with a temporal change detector, their input functions grow and decay over time. When this growth function is used to excite the Gaussian activity pattern, one obtains a nxed-scale Gaussian activity wave with amplitude that grows and then decays in time. Grossberg and Rudd demonstrate that if spatially separate inputs are flashed at different

times, for an appropriate scale Gaussian the two activity waves will merge into a single activity hump, its maximum sliding continuously from the initial to the final input. They then assume a separate contrast-enhancement (CE) process localizes this moving maximum.

The DEB model shares the two essential elements of the Grossberg-Rudd model, i.e., a spatially extended response to an input that evolves over time, followed by a CE process that localizes the response; however, where Grossberg and Rudd assume a fixed-scale Gaussian response to an input, the DEB utilizes a diffusion process that responds with increasing scale as a function of time.

# 3. PSYCHOPHYSICAL RESULTS

## 3.1 Spatiotemporal Grouping of Static Images

The DEB model, while most closely associated with grouping in dynamic images, also replicates some effects of grouping in static images. A striking example of visual static feature grouping is demonstrated by Marroquin's diagram shown in Figure 1 [41]. An observer, staring at the center of the hexagon, will see dots appear to group with their neighbors on increasingly greater spatial scales as a function of time. Although the image is static, the grouping that takes place seems to be a dynamic process with larger scales requiring more time to emerge.

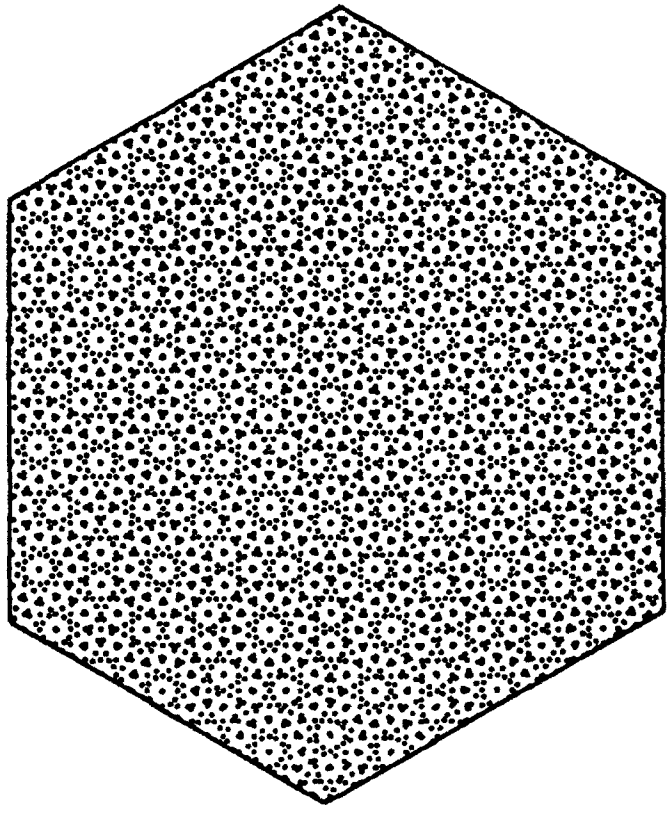

*Figure 1. Static feature grouping on multiple scales; evidence for the existence of a dynamic grouping process in the visual system [41]. Reprinted with the permission of MIT.*

### 3.2 Spatiotemporal Grouping in Dynamic Presentation

The simplest dynamic apparent motion effect, gamma motion, occurs when a single light is turned on briefly and turned off. Although the light is a fixed spatial extent, the perception is a light that first expands and then contracts [42,4], as shown in Figure 2 for 1-D space. The expansion is "more impressive" than the contraction, especially at longer presentation times, when the contraction is "appreciably weaker and less extensive" [42]. This expansion also occurs when a dark spot is introduced on a uniform field, and the contraction is perceived when the spot is removed. The effect is identical regardless of direction of contrast, thus suggesting it occurs in later visual motion processing (long-range or second-order system).

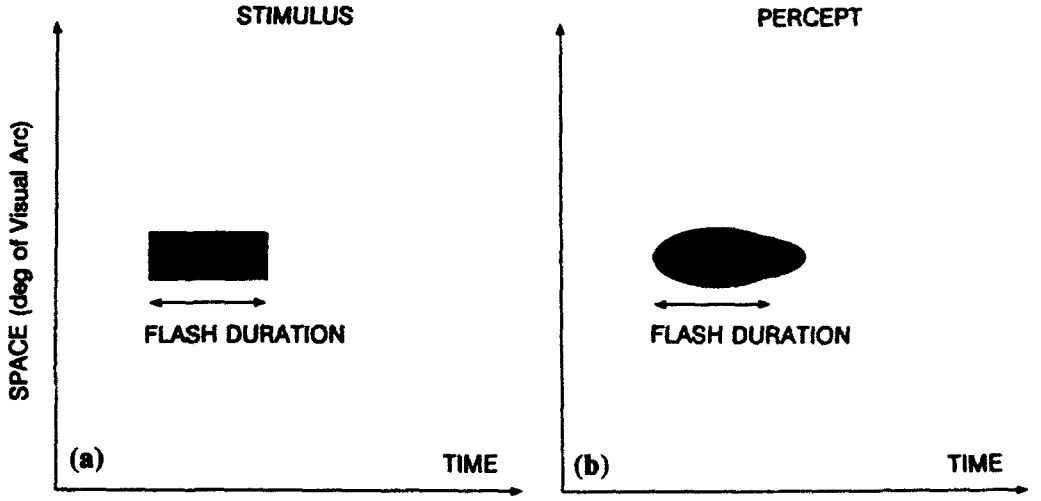


*Figure 2. Gamma motion. (a) A light of fixed spatial extent is illuminated then extinguished. (b) The percept is of a light expanding and then contracting. For long flash durations the contraction percept is noticeably weaker and less extensive than the expansion percept.*

New dynamic grouping phenomena emerge when two distinct stimuli interact over time to form the percept of long-range apparent motion. In the human visual system apparent motion can be demonstrated with two lights of fixed spatial extent that are illuminated at distinct times across a fixed spatial separation (Figure 3). With different spatial separations, illumination times, and interstimulus intervals, the percept can appear as two separate lights flashing; as one spot that moves smoothly between two real lights; as one spot that moves smoothly from the first location, jumps, and continues moving smoothly to the second location; or as two unrelated spots [4,5].

Similar apparent motion effects can be achieved with tones in the auditory system moving across space [7] or pitch [6].

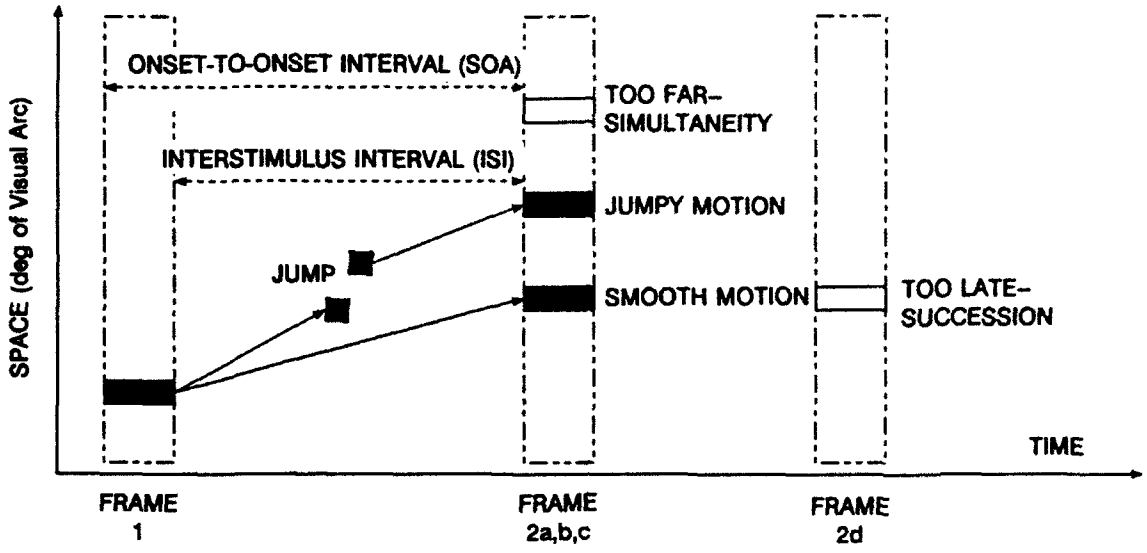


*Figure 3. Long-range apparent motion. Filled rectangles represent sources in space-time separations that produce the illusion of long-range apparent motion. Empty rectangles represent sources that are ignited too late for the spatial distance or are too far away for the given ISI to produce the sensation of motion. Jumpy motion rectangle occurs soon enough to give the appearance of smooth motion from the leftmost rectangle to the first shaded square, a short jump to the second shaded square, followed by smooth motion to the destination, while the closer rectangle below it exhibits pure smooth motion.*

The most common form of object motion, phi motion, gives no impression of a particular shape undergoing movement, whereas in beta motion a well-defined shape is seen. Kolers [4] performed a series of experiments that pitted long-range apparent motion between like shapes against long-range apparent motion of patterns. In these experiments filled circles could move to filled circles and filled squares could move to filled squares, or the group of objects could move en masse with circles moving to squares and vice-versa. After systematically examining these displays, he concluded that over a wide range of ISI, individual figural identity is subordinate to the global stimulus pattern. Ullman [39] developed additional displays and likewise concluded that "no elaborate form analysis must precede the correspondence operation." Thus in the visual domain, it is believed that the visual array is minimally preprocessed before being input to the DEB network.

Psychologists have examined in detail the stimulus presentation conditions that produce simultaneity, smooth motion, jumpy motion, or succession and discovered that for fixed-flash durations there is a clear range of onset-to-onset interval, often referred to as the "stimulus onset asynchrony" (SOA)[2] versus spatial separation that will produce smooth apparent motion (Figure 4). If one shortens the SOA, lights begin jumping rather than smoothly moving between each other, while a still shorter SOA causes the lights to appear to flash simultaneously. If one lengthens the SOA beyond an acceptable limit, the lights flash independently. Similar conditions can be created by fixing the SOA and varying the spatial separation of the two stimuli.

Reverse motion can occur when the duration of the first stimulus is much larger than that of the second. In this case motion is generated from the first light to the second light as previously stated but then springs back to the first light without a new presentation of the first stimulus [4].

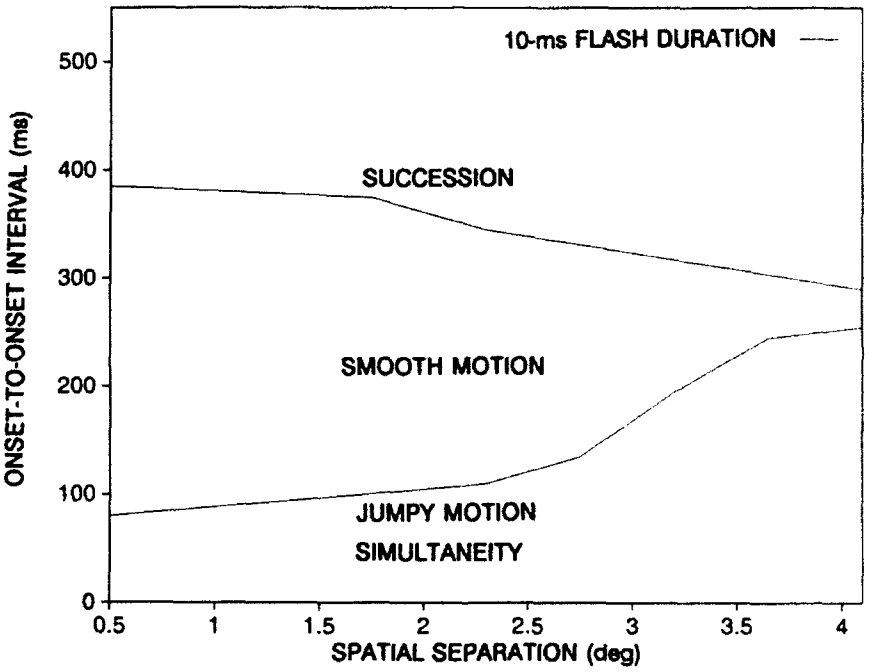


*Figure 4. Retinal long-range apparent motion regimes. SOA versus spatial separation for 10-ms flash. The jumpy motion region is not quantitatively delineated. (Plot after Kolers [4], using Neuhaus data [43].)*

[2]Defined as the time between the onset of two successive applications of a stimulus. Related is the ISI that is defined as the time between two successive stimuli. Thus ISI + stimulus duration = SOA.

Once phi motion was discovered, the next step was to calculate its apparent velocity. Unfortunately, it is unclear how best to calculate velocity of the illusory motion. A physical theory would plot $V = \frac{Distance}{Time}$; in these experiments the interpretation of "distance" and "time" is not clear. Distance in degrees of visual angle or retinal distance is a natural choice, but the spatiotemporal grouping occurs in cortex, after a logarithmic spatial compression. If ISI is used for time, cases exist for illusory motion with infinite velocity. SOA is the next reasonable choice, but it has been shown that stimulus duration affects apparent motion. Kolers [4] argues for retinal distance and SOA; a variant of his graph is displayed as Figure 5 to aid comparison with the results of the current report simulation. It is unclear, however, whether the relationship between velocity and spatial separation is linear in long-range apparent motion. (Although it is never clear that the ratio of velocity to spatial separation is linear, longer flash durations are closer to linear than the 10-ms flash duration plotted here.) This unusual velocity behavior is quite distinct from that experienced in the short-range system.

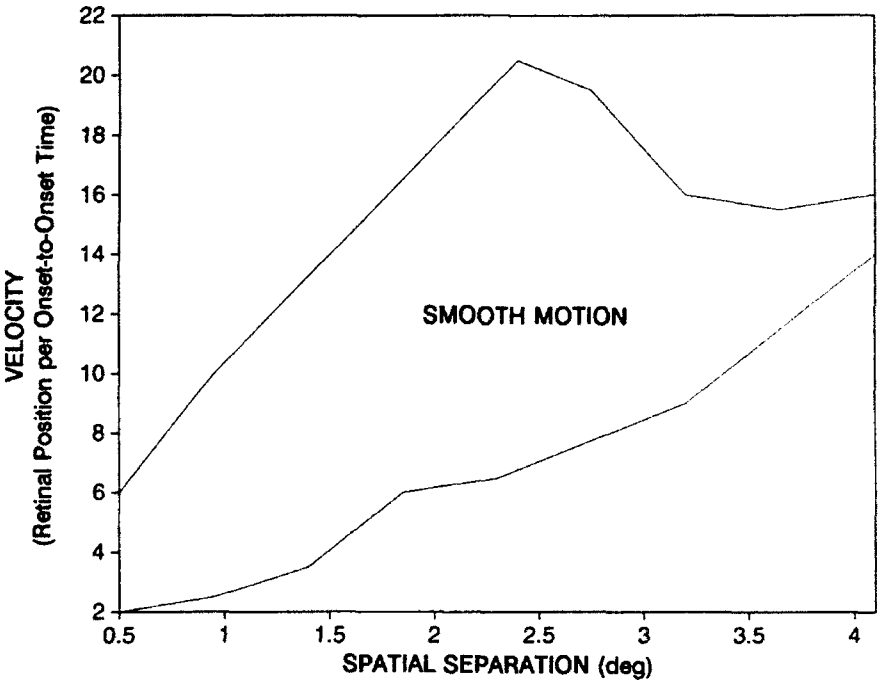


*Figure 5. Calculated velocity versus spatial separation in observed long-range apparent motion for a 10-ms flash. (Plot after Kolers [4], using Neuhaus data [43].)*

To further understand apparent motion, other psychophysical experiments have been performed; in the visual domain all involve additional stimuli. Merge motion effects are demonstrated with three lights of fixed spatial extent. Two outer lights are illuminated and extinguished, followed by illuminating and extinguishing a central light. If the two outer lights are equidistant to the center light, the former appears to merge and move to the center light. If the two outer lights are at staggered distances, only the closest light moves to the center light (see Figure 6). If the equidistant display is placed in the periphery of the visual field, the outermost light appears to move toward the center.

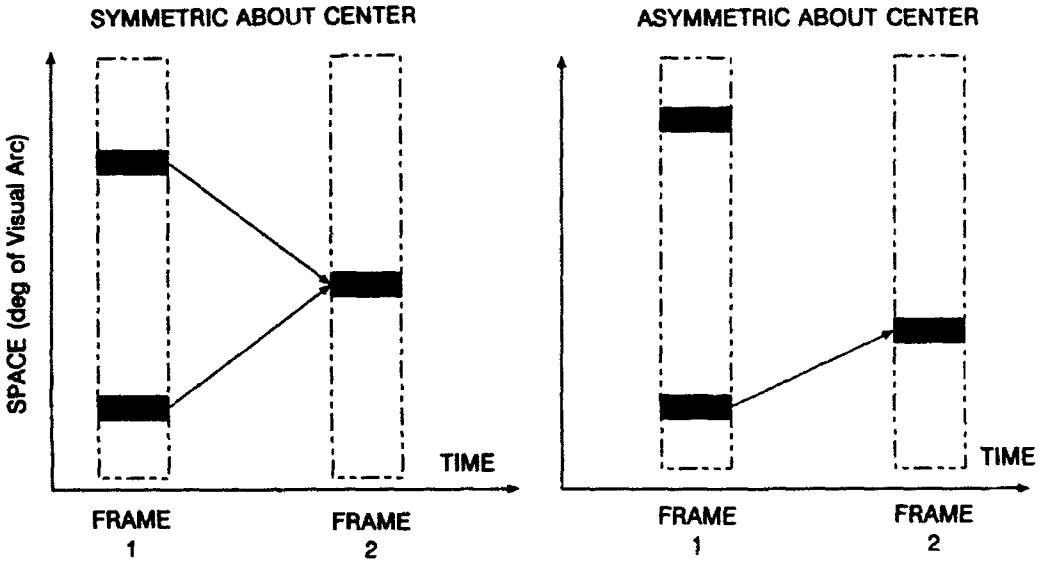


*Figure 6. Merge motions. Illuminating two point sources equidistant from a third, later-illuminated point source gives the appearance that the two merge and become one. If the original pair is not equidistant from the third point source, only the closer of the two will appear to move. Alternatively, if the symmetric case is placed in the periphery, only the most peripheral light will move to the center.*

The opposite of the merge motion effect is split motion, in which first the center light, then the two outer lights are illuminated and extinguished. If the two outer lights are equidistant to the center, they both appear to move to the center light and merge with it. If the two outer lights are not equidistant, the center light appears to move to the closer (Figure 7). Once again, if the equidistant version of the display is placed in the periphery of the visual field, the center light appears to move toward the light farthest from the fovea.

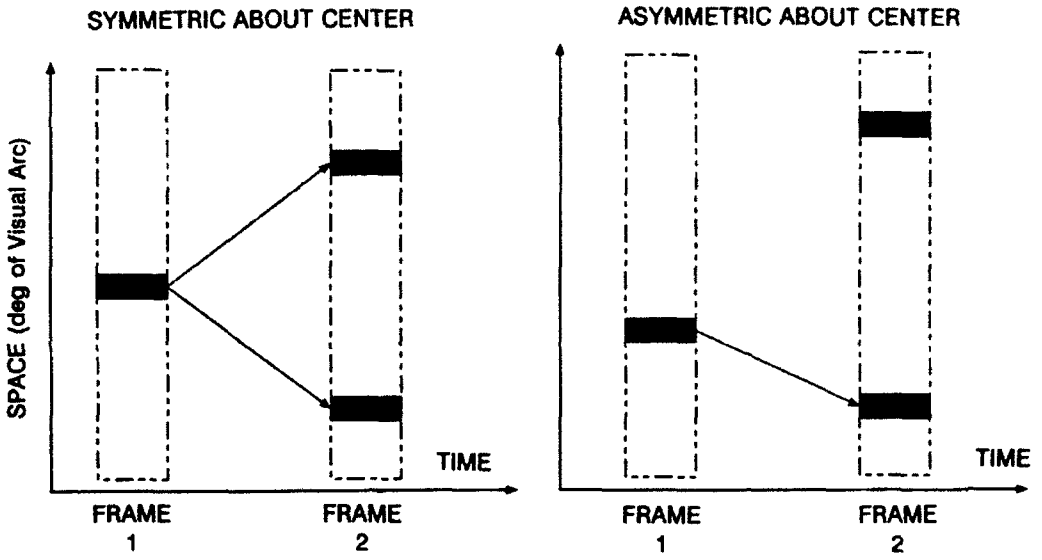


*Figure 7. Split motions. Illuminating one point source followed by illuminating two equidistant point sources cause the first to appear to split and move to both later sources. If these are not equidistant to the first, movement is only to the closer of the pair. If the symmetric case is placed in the periphery, motion will be from the center to the most peripheral source.*

A variation of the above two-frame displays, called dichoptic presentation, presents the first frame to the subject's left eye and the second frame to the subject's right eye. All the two-frame displays result in the same percept when dichoptically presented.

Another important multielement stimulus, originally developed by Ternus [44], illustrates that more than one distinct motion percept can be produced by the same display. In the Ternus display, two frames with three lights each are illuminated and extinguished in succession. The frames are aligned so that two of the three lights occupy the same space, and the third appears alternately on the left and right of the central two objects. When the ISI is short, the third light appears to move around the central two objects (element motion), but when the ISI is long, the three lights appear to shift as a coherent group (Figure 8) [4,23]. Both types of motion occur even when the two center lights are not perfectly aligned across frames [45], but larger perturbations favor group motion. Group motion is also favored when stimulus contrast is reduced or background luminance during the ISI is increased above stimulus presentation luminance [23]. Finally, dichoptic presentation or stimulus contrast reversal always results in group motion [46]. These variations suggest that group motion over large separations may be due to a separate process distinct from the short-range system. In addition, the percept is never a mixture of element and group motion, thus strengthening the argument for multiple motion systems [23].

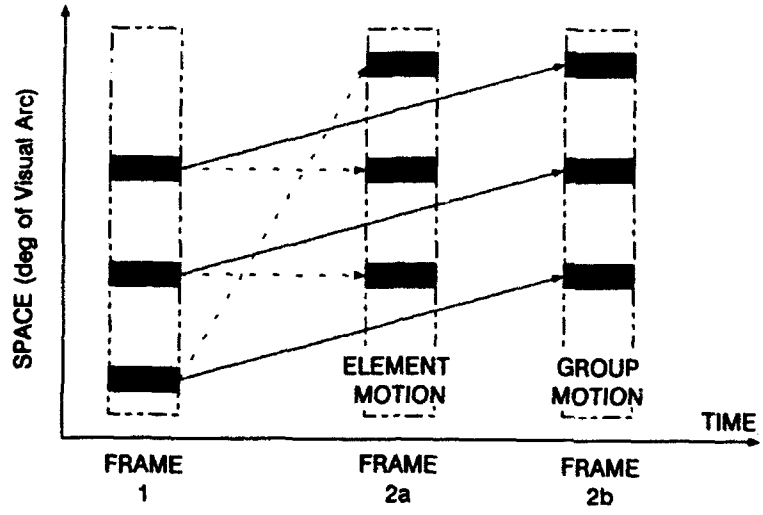


*Figure 8. Ternus motions. Illuminating shifting groups of three point sources produces the illusion of one in element motion or all three in group motion. The percept changes with changing ISI.*

This spatiotemporal grouping process is not restricted to the visual system. In the somatosensory system, if two vibrators agitate the skin with a small ISI, the subject experiences a single vibration between them. As in the visual domain, the effect originates not in the skin but in the cortex; if the skin between the two vibrators is locally anesthetized, the effect is still experienced [8]. Similar experiments have been devised for the auditory system across space [7] and pitch [6], and similar results were reported. Most bizarre are the intermodal experiments in which apparent motion is perceived between a sound source and a light source [7].

### 3.3 Useful Psychophysical Parameters

Psychophysical literature provides evidence for spatiotemporal grouping in long-range apparent motion and suggests rates at which it should occur. Figure 4 provides spatial and temporal parameters to which the model should conform. The lower curve suggests that communication time varies with distance (i.e., Korte's law), while the upper curve suggests that stimulus memory (and its reinforcement mechanisms) fades with time. Furthermore, this fade occurs faster with increasing distance. Both are consistent with a leaky diffusion network in which stimulus activity first rises and spatially expands, then falls and contracts; the DEB network was designed to produce these effects. For a fixed spatial separation, there should be a range of SOAs in which motion should be perceived, and outside that range no motion should be realized.

It is difficult to directly relate the time/space scales of the DEB to biological time/space, so ratios are examined that allow units to be ignored. One ratio considered is that of the longest to the shortest SOA for which smooth motion occurs at a fixed spatial separation. Figure 4 suggests that this ratio is restricted to $\leq 4$ with the ratio decreasing as the spatial separation increases. It is not clear how to best compare psychophysical with model-predicted velocities, because there is still no way to interpret the perceived velocities of apparent motion. Nevertheless, in Section 6.4.4 the DEB simulations will be compared to Kolers' definition of psychophysical velocities.

# 4. BIOLOGICAL CONSIDERATIONS

## 4.1 Cortical Role and Pathways

Consider that the preceding psychophysical effects have two salient processes: long-range communication that facilitates interaction of features generated by the inputs (flashing point light sources or sinks in the visual domain) and focusing that enables stimuli in apparent motion to have a definitive location. This section discusses the visual neurobiological elements that may provide the foundations for both processes.

Although the cause of these long-range motion effects has been debated for nearly a century, it is known that they do not occur at the retinal level as is evident from the dichoptic long-range apparent motion experiment. In that experiment apparent motion is experienced, indicating that spatiotemporal interactions do occur at the cortical level. In addition, and based on the split and merge experiments performed in the periphery, these effects occur after the visual system has compressed the periphery in favor of the fovea, i.e., at the visual cortex. Such sampling and compression occur in at least two places in the visual system [47,48]. First, the retina itself contains a nonuniform population of rods and cones; the densest population occurs in the central fovea. Second, the receptive fields of the retinal ganglion cells are markedly smaller in the fovea than in the periphery, hence, the well-known cortical magnification of the fovea and compression of the periphery. Based on neurobiology and psychophysics, it seems evident that the substrate of apparent motion lies in the visual cortex.

The cortex is divided almost evenly between nonneuronal and neuronal cells. We suggest that the nonneuronal astrocyte glia may form the substrate of the long-range communication process, while neuronal networks provide focusing to create and reinforce the localized percept.

There are strong indications from psychological, neurophysiological, and perceptual literature that two rather separate visual pathways exist; one that primarily responds to movement and disparity, and one that responds to static attributes such as form and color [49–51]. The dynamic pathway is commonly referred to as "magnocellular." In primates, the anatomical and physiological differentiation from the static or parvocellular pathway is clear as early as the retinal ganglion cells. Large, type-A retinal ganglion cells provide input to the large-celled magnocellular layers of the lateral geniculate body, while smaller, type-B cells provide input to its parvocellular subdivision. From here the magnocellular pathway progresses up through visual cortex area V1, layers 4C$\alpha$ into 4B; then into area V5 (MT) both directly and via areas V2K and V3. The parvocellular pathway enters area V1, layer 4A both directly and via 4C$\beta$, then enters layers 2 and 3 before connecting with areas V2I and V2N, which in turn connect to V4. After this point connections are more broadly made and less well understood. One possibility is that the direct V1-V5 connection is concerned with first-order, short-range processing, while the pathways via V2 are associated with second-order, short-range processing, and the V1-V3-V5 pathway (if it includes neural-astrocyte glial interactions, in which the astrocytes provide a long-range signaling mechanism) could perform

long-range processing. The two visual pathways (magnocellular and parvocellular) permit a similar network to be used in each for spatiotemporal grouping of time-varying or static inputs.

### 4.2 Astrocyte Glial Cells

Once thought of as only providing passive physical support, neuroglial cells now appear to play an "active role in maintaining normal brain physiology" [52]. The astrocyte glial cells are concentrated on because they are known to provide long-range communication between coupled astrocytes [52]. Although to date such coupling has not been directly demonstrated in vivo (Kettenman and Ransom [53] suggest that this is due to technical difficulties), there is some evidence that it occurs,[3] and there is direct evidence for coupling in cultured astrocyte cells. Such communication is not rare. Indeed, Kettenman and others have observed that "mammalian astrocytes in cell culture are widely coupled to one another electrically" and that "qualitative studies have shown that cultured astrocytes form a highly coupled electrical syncytium" [53] which is postulated as providing the long-range communication necessary to support the above psychophysical phenomena. A diffusion layer (comprising a network of local currents) is an essential part of the current DEB network.

An alternative long-range astrocyte glial signaling mechanism has also been recently suggested [57]. In this experiment, glutamate, a common excitatory neurotransmitter, induces propagating waves of $Ca_i^{2+}$ in cultured astrocytes. It is possible that these waves induce later neuronal activity. Either of these mechanisms may be responsible for the long-range interactions simulated here.

### 4.3 Neuronal Networks

Many cells in the visual cortex are known to derive their input from networks of neurons. In Hubel [47], simple cells (which respond to oriented lines) are postulated to be made up of a hierarchy of lower-order, radially symmetric, center-surround cells. Similarly, complex cells (which respond to oriented lines in a wide receptive field) and end-stopped cells are made up of a network of simple cells. Also, directionally tuned, motion-sensitive cells are postulated to consist of inhibitory and excitatory connections. Similar lateral inhibitory and self-excitatory interconnections have been used by Grossberg [13] in shunting neural network architectures that contrast-enhance their inputs. The DEB network also utilizes a CE network to localize the stimulus percept and provide a reinforcing feedback signal.

---

[3]Low molecular weight dye passes between adjacent cells [54], and glial networks are postulated to act as potassium spatial buffers [52,53,55,56] for nearby neurons.

### 4.4 Interactions

Astrocyte glial cells are known to interact electrically among themselves and with neurons as well; the nature of that interaction is now considered.

Kettenman and Ransom [53] discovered that in cultured astrocyte syncytia, the resistance of the electrical gap junctions between astrocytes is not voltage-dependent over much of the membrane potential fluctuation, thus a charge flow model examining interastrocyte communication should be independent of the membrane potential of the astrocytes. Furthermore, it is known that "glial cells...have a high potassium concentration and have negligible ionic permeability for ions other than potassium" [55]; therefore, current flow is modeled between glial cells as transfer of potassium ions to or from the cell in a manner obeying Ohm's law. When this process is expanded to encompass current flow in a network of glial cells, the motion of these ions can be approximated with a diffusion equation.

It is further proposed that a glial syncytia provides long-range communication between neurons in a layer via transmission of potassium and other ions. It has been shown that "variations in $[K^+]_{extracellular}$ have profound effects on neuronal excitability, modulating such processes as synaptic transmission and the initiation and propagation of action potentials" [56]. Such $[K^+]$ variations can be realized near the leaky endfeet of astrocytes that are in close proximity to neuronal synapses. This report is not the first to propose such an interaction. In 1965 Hertz [58] proposed "a mechanism...in which the potassium ions, which have been lost from one nerve cell during its activity, are transported through neuroglia cells to the outer surface of another nerve cell, which is then depolarized and stimulated; that is, a neuronal-neuroglial-neuronal impulse transmission." Hertz continues: "Potassium ions which have been released from an active area are transported through neuroglia cells to the outside of other neurones; these are in turn stimulated and potassium ions are released, to be transported actively through other neuroglia cells. In this way the spreading depression is propagated across the entire cortex more rapidly than can be explained by a diffusion." An alternative rapid propagation mechanism could be due to small interglial electromagnetic fields, where ions taken up at one location induce other ions to flow. The DEB model explicitly uses such interactions to spread and reinforce the charge distribution in a diffusion layer.

### 4.5 Useful Biological Parameters

Odette and Newman [56] note that glial cell endfeet "can contain up to 95% of the total cell conductance." This information is important to determine how "leaky" the diffusion process should be.

Kettenman and Ransom [53] have examined astrocyte coupling in cultured layers by electrically stimulating (via KCl injections) one glial cell and measuring its voltage and that of the neighboring cell. The ratio of these voltages are fit to an exponential, which approximates the steady-state decay in a 1-D and 2-D syncytium: $\frac{V_2}{V_1} = \exp\left(\frac{-d}{L}\right)$, where $d$ is the distance from the injection, and $L$ is a decay-length constant. Kettenman and Ransom [53] measured astrocyte $L$ in

vitro to be 80 to 100 $\mu$m. $L$ can be used to relate the DEB model to physical size of the biological networks and is related to the ratio of the conductances given by $G_{gg}$ and $G_g$ in Section 5.1. Further, $L$ is not related to the 1-D model explained herein, as it is believed that the 1-D length constant would have to be significantly greater than the 2-D decay length. Indeed, experiments with restricted 2-D syncytia have $L$ values that are greater than their full 2-D syncytia counterparts [53]. The measured value for $L$ is expected to be more useful in the context of 2-D experiments.

One critical missing piece of data is the *rate in which activation due to $K^+$ influx at one glial endfoot affects $K^+$ efflux at another.* Measurements indicating the time for low molecular weight dyes to pass through an in vitro cell are available, but this is not the desired measurement [54]. Such measurements are central to the hypothesis that an astrocyte glial layer may act as a long-range communication network supporting the spreading of activity among neighboring neurons.

# 5. DIFFUSION-ENHANCEMENT BILAYER MODEL

## 5.1 DEB Network Architecture

The DEB model consists of two processes that mirror the two salient psychophysical processes mentioned in Section 4, i.e., a diffusion layer that facilitates long-range interactions via local connections, and a CE layer that focuses and reinforces the diffusion layer and provides the sensation of a localized object traversing a spatial separation. In the DEB model, visual input is preprocessed to indicate the location of the pattern without indicating its shape or direction of contrast. This method is in keeping with the relative strength of phi versus beta motion as indicated by Kolers' shape experiments. This preprocessed input is passed to the spatiotemporal grouping network. In the case of the primate vision system, both center-surround processing and logarithmic spatial mapping occur before grouping begins in the cortex [47].

Following feature extraction, activity is input to a leaky diffusion layer where it spreads and interacts with a localizing CE layer, which periodically samples the state of the diffusion layer. Output from the CE layer is fed back to the diffusion layer to reinforce new input and facilitate sustained interactions. This report proposes that a short-range motion detection system detects the smooth motion of the activity maximum as well as the motion of the activity edges (it is equally valid to detect any activity contour line) at the output of the CE layer and causes the sensation of motion in the psychophysical experiments. Also, activity prompted by a single input at first grows and eventually dies down, so that after a period of time grouping is no longer possible (Figure 4). This is a combined effect of the limited time span of featural input from a single feature, the leaky diffusion layer, and the imposition of decay on the feedback from the CE layer.

With this high-level description of the network in mind, a 1-D circuit form of the DEB model is illustrated in Figure 9. Note the two layers—a diffusion layer that permits long-range charge interactions and a CE layer that localizes the charge distribution from the diffusion layer and produces improved signal-to-noise ratio via the feedback pathways. A separate input layer of feature-sensitive neurons (e.g., transient response ganglion cells) provides activity to the diffusion layer via glial cell endfeet. The form of this input is left undefined until Section 6.2 because it is expected to change between sensory systems and pathways. Glial endfeet also support feedforward and feedback activity flow (i.e., charge or $K^+$ ions) between the diffusion and CE layers.

## 5.2 Diffusion Layer Equations

The diffusion layer of the DEB model is governed by three coupled systems of differential equations based on Ohm's law; the first represents the 1-D spatially coupled nodes:

$$\frac{dQ_g^{(i)}}{dt} = \frac{G_{gg}}{C_g}\left[Q_g^{(i+1)} + Q_g^{(i-1)} - 2Q_g^{(i)}\right] - \frac{G_g}{C_g}Q_g^{(i)} + \frac{G_{ge}}{C_g}\left[\frac{C_g}{C_e}Q_e^{(i)} - Q_g^{(i)}\right] + \frac{G_{gi}}{C_g}\left[\frac{C_g}{C_i}Q_i^{(i)} - Q_g^{(i)}\right] \quad . \tag{1}$$

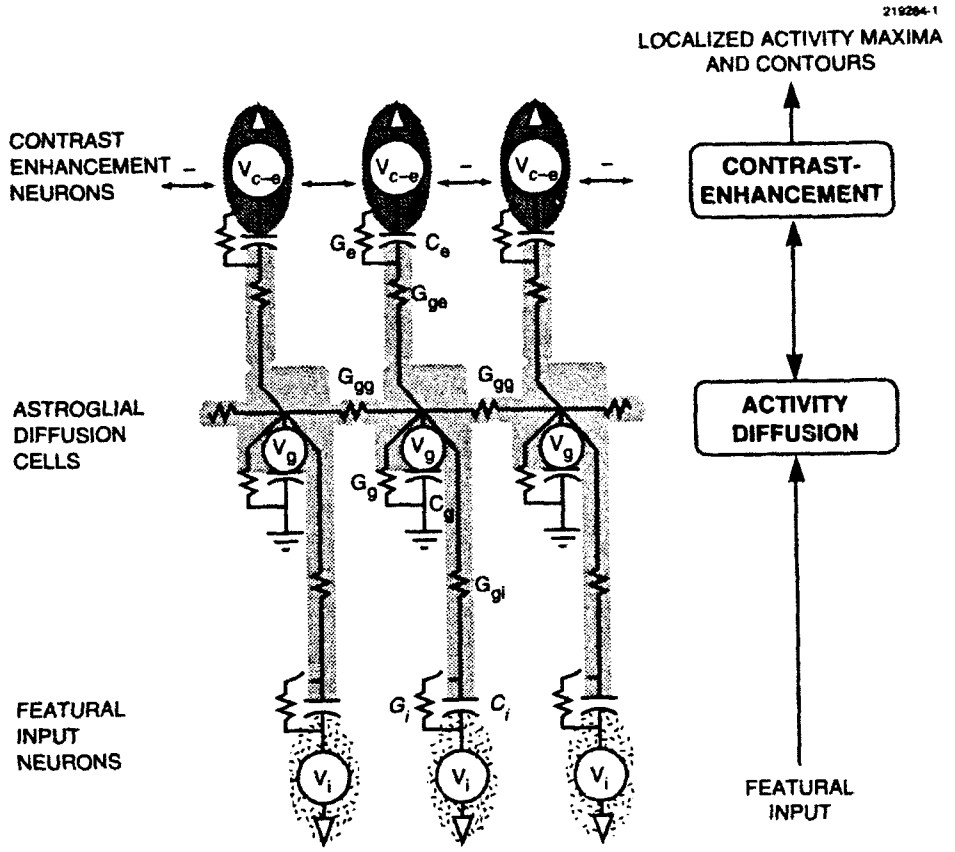


*Figure 9. DEB network architecture. Charge is distributed throughout the network via the dynamics of Equations (1) through (8). Note the two conceptual layers (and the input) that provide long-range communication (astroglial diffusion cells) and localization (CE neurons). Bidirectional interfaces between layers are formed by glial endfeet.*

Equation (1) contains several parameters that can be considered independent of the other coupled equations. Conductivity $G_{gg}$ controls the speed with which charge $Q_g$ is distributed throughout

the glial layer, while $G_g$ controls how rapidly charge leaks from the glial nodes into the environment. Together, conductivities $G_{gg}$ and $G_g$ determine the spatial extent over which charge can spread in the diffusion layer. Equation (2) governs the rate at which featural input ($I^{(i)}(t)$) enters the diffusion layer via glial endfeet, while Equation (3) governs the rate at which feedback from the CE neurons ($F^{(i)}(t)$) enters via the CE endfeet:

$$\frac{dQ_i^{(i)}}{dt} = -\frac{G_{gi}}{C_g}\left[\frac{C_g}{C_i}Q_i^{(i)} - Q_g^{(i)}\right] + I^{(i)}(t) \quad , \tag{2}$$

$$\frac{dQ_e^{(i)}}{dt} = -\frac{G_{ge}}{C_g}\left[\frac{C_g}{C_e}Q_e^{(i)} - Q_g^{(i)}\right] + F^{(i)}(t) \quad . \tag{3}$$

Conductance $G_{gi}$ controls the rate at which new inputs affect the charge profile on the diffusion layer. $G_{ge}$ controls the rate at which the CE layer feels the effects of developing charge distributions on the diffusion layer as well as the rate at which feedback from the CE layer modifies the diffusion layer. The capacitors represented in all three equations store the distribution of charge in the diffusion layer ($C_g$), the interfaces (endfeet) to the input ($C_i$), and the enhancement layer ($C_e$).

### 5.3 Contrast-Enhancement Layer Equations

The charge in the CE endfeet is periodically sampled by the CE neurons and can be identified with the refractory period of neurons that are phase-locked in a layer; CE neurons process activity on a shorter time scale than the diffusion layer. The sampled charge is contrast-enhanced via a network originally formulated and analyzed by Grossberg [13], and the output from this network is fed back to the facing endfeet. The equation governing charge in this system of $N$ neurons represents a network of self-exciting nodes with long-range lateral inhibition (over the entire layer) and passive decay;

$$\lambda\frac{dQ_{ce}^{(i)}}{dt} = -AQ_{ce}^{(i)} - Q_{ce}^{(i)}\sum_k f(Q_{ce}^{(k)}) + Bf(Q_{ce}^{(i)}) \quad , \tag{4}$$

where

$$f(Q) = \begin{cases} CQ^2/Q_1 & 0 \le Q \le Q_1 \\ CQ & Q_1 < Q \end{cases} \quad . \tag{5}$$

Equation (4) can be rewritten as a shunting short-term memory model with charge limited to the range $[0, B]$. Depending on the choice of parameters, the rapidly attained equilibrium can

either pick the node with maximal charge or contrast-enhance the charge across the layer. The latter properties are of interest here where constant signals are suppressed, noise fluctuations are quenched, and all nodes of nearly maximal activity are enhanced. In any case, the dynamics lead to a normalization of activity across the layer, with total equilibrium activity $E = B - \frac{A}{C}$. When in this domain, the nodes for which activities fall below

$$Q(t) = \frac{Q_1}{B - \frac{A}{C}} \tag{6}$$

for a sufficiently large time will be forced to lose all activation, i.e., they will be quenched. After Equation (4) equilibrates, the amount of feedback in (3) is $F^{(i)}(t) = Q_{ce}^{(i)}$

Because feedback reactivates the diffusion layer, even once the original input is off, there is the need to dampen the feedback amplitude over time. Without this step a single light will be sustained in memory forever. This problem is resolved by forcing the parameter $B$ in Equation (4) to decay with time between inputs to the system. That is, the duration of the feedback is limited. When a new input stimulates the visual field, a delayed signal reenergizes the CE layer and $B$ is reset to its maximum value. Specifically, $B$ is modeled as

$$B(\hat{t}) = \left[ B_{max} - \left( \frac{\hat{t}}{\tau} \right)^2 \right]^+ \quad , \tag{7}$$

where

$$\hat{t} = t - t_{most\ recent\ stimulus} - t_{delay} \quad , \tag{8}$$

and where $[x]^+$ is a threshold linear function equal to $x$, if $x \geq 0$, and 0 otherwise.

### 5.4 Relationship to Biological Networks

A number of DEB model components have direct correlates to biology. Capacitor $C_i$ represents a neuronal-astroglial interconnect that is locally excited by presented features. A feature-sensitive neuron fires; as it repolarizes $K^+$ is released into the extracellular compartment. (The input source function described in Section 6.2 tries to model this charge release.) $K^+$ ion pumps bring $K^+$ onto $C_i$, which represents a highly permeable endfoot of an astrocyte glial cell (cf. $K^+$-spatial buffering [55,56]). Once inside, the $K^+$ is freely diffused via ion currents within glia (with membrane capacitance $C_g$) and interacts with other cells through electrical gap junctions between astrocyte glial cells [53]. The interglial connections are represented by the conductors $G_{gg}$. A portion of the $K^+$ is diffused out of the cells at endfeet to an upper, contrast-enhanced neuronal layer, which is excited by the locally increased extracellular $K^+$ concentration. This astroglial-neuronal

interconnect is represented by $C_e$. It is hypothesized that the neuronal layer interacts within itself to contrast-enhance its own activity, further releasing $K^+$ as it fires. This CE layer, triggered by the excess $K^+$ released by nearby glia, amplifies activity via a flow of other extracellular ions into the neurons. The extracellular driving potential for these ions is represented by the parameter $B$ in Equation (4). As these neurons fire, the extracellular ion concentration drops due to its consumption, and is modeled by the decay of $B$ according to Equation (7). The resulting contrast-enhanced $K^+$ profile is fed back to the glial layer via the same endfeet, and thereby reinforces the charge distribution in the glial network, particularly near the charge maximum. The output of the CE layer also provides the basis for the percept of a compact form in smooth motion by exciting a short-range motion system.

# 6. NUMERICAL SIMULATIONS

## 6.1 Overview

The following numerical simulations illustrate the dynamics of the DEB network in the context of spatiotemporal grouping and replicate many of the psychophysical percepts described previously. In all the following simulations, the model parameters are set to the parameter values illustrated in Table 1. Furthermore, except where indicated, a small amount of noise is added to $C_i$, $C_g$, and $C_e$. The total noise per node is randomly distributed between 0 and 1 unit of activity. This simulates the effects of additive noise that could be caused by residual activity or steady-state random-onset firing of the input neurons. Without the ability to suppress such noise, feedback would generate false sources of activity by amplifying endogenous noise [2]. By causing the total feedback activity to slowly increase or by delaying the onset of feedback after the stimuli's presence is felt, more recent stimuli are able to enter the network and overwhelm the noise.

**TABLE 1**

**Simulation Parameter Values**

| Parameter | Dynamic | Static |
|---|---|---|
| $\frac{G_g}{G_{gg}}$ | 0.25 | 0.25 |
| $\frac{G_{ge}}{G_{gg}}$ | 0.4 | 0.4 |
| $\frac{G_{gi}}{G_{gg}}$ | 0.003 | 0.7 |
| $\frac{C_g}{C_e}$ | 15 | 15 |
| $\frac{C_g}{C_i}$ | 10 | 10 |
| $Q_1$ | 23 | 23 |
| $A$ | 1 | 1 |
| $B$ | 2501 | 2501 |
| $C$ | 1 | 1 |
| $t_{delay}$ | 2 | 0 |
| $\tau$ | 2.25 | 2.25 |

In all simulations, the motion of activity edges and maxima within the CE layer is considered to drive the percept of apparent motion (by exciting a short-range motion process) with the maxima defining object location.

For each simulation several graphs are plotted, showing the evolution of activity across four levels of the network: at the input endfeet ($C_i$), the interglial diffusion site ($C_g$), the enhancement endfeet ($C_e$), and the output of the CE network. On the graphs labeled "Enhancement Endfeet," contours start at 0 and are drawn every 5 units up to 45 units of activity, while on the other graphs the contours depict activity starting at 0 and are drawn every 50 units up to 450. A connected line depicting the trajectory of the local maxima is superimposed on the contour plot illustrating the output of the CE layer.[4] Local maxima are determined at 5× higher resolution than the activity plots presented; down-sampling was required by the printing process. Thus rising and falling activity waves are smoother than displayed.

### 6.2 System Input

Two different inputs are used to model the result of early visual processing done in each of the two visual pathways. The first is a sustained input used in the static pathway, which rises rapidly to a maximum value that is then sustained until the stimuli are removed. The second is a transient input used in the dynamic pathway, which rises rapidly then decays quickly. Although attempts were made to produce results consistent with the psychophysical database for static and dynamic stimuli using one version of the model and one type of input, these attempts failed. Exploiting the neurophysiological literature, it was noted that cat $X$ and $Y$ cells and primate $A$ and $B$ cells had significantly different stimulus response characteristics [59,60]. A functional approximation to the "average response histogram" recorded by Enroth-Cugell and Robson [59] was then generated. The two inputs are shown in Figure 10. The static input is defined as

$$I(t) = \begin{cases} 20 * t^{1.1} * e^{\frac{-t}{15}} & 0 \leq t < 15 \\ 20 * 15^{1.1} * e^{\frac{-15}{15}} & 15 \leq t < 100 \\ 0 & \text{otherwise} \end{cases} . \tag{9}$$

while the dynamic input is defined as

$$I(t) = \begin{cases} 26 * t^{1.1} * e^{\frac{-t}{20}} & 0 \leq t < 70 \\ 0 & \text{otherwise} \end{cases} . \tag{10}$$

[4] In some cases, several additional extremely weak local maxima also exist, but because they are more than 5 orders of magnitude smaller than the depicted local maxima, it is assumed that limited precision neurons would be unable to "sense" them.

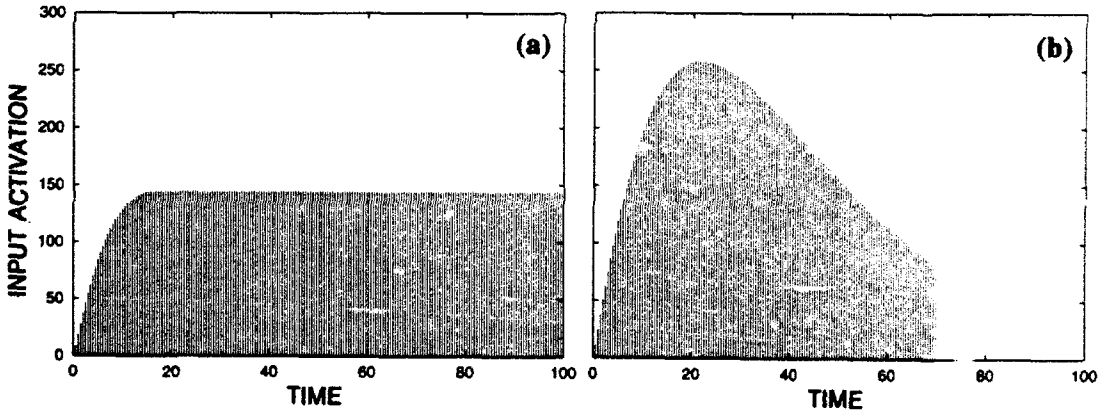


*Figure 10. Simulation inputs. (a) Sustained input to the parvocellular system; (b) transient input to the magnocellular system.*

These inputs are applied as point sources to the input endfeet $Q_i$ at node $i$, as indicated by $Q_i^{(i)}$ in Equation (2). In this report the inputs are not modulated according to stimulus intensity effects (cf. Korte's law [4]), in part because the obvious definition of stimulus luminance is often confounded with stimulus size and figural detail. A natural extension to the current work, however, would link stimulus luminance with input activity.

In the simulations presented here, only positive transients (ON cells) are input to current networks, but it is believed that both positive and negative transients interact in a single DEB network. Preliminary simulations suggest that similar results are obtained but that activity dynamics occur over longer time scales.

All simulations begin at $t = 0$ with one or more inputs, and in the dynamic simulations additional inputs are added at later times. Equations (1) through (3) are integrated with a Runge-Kutta integrator. Periodically (every time unit), $Q_e$ is sampled and used as the initial conditions for Equation (4), which is integrated with a Runge-Kutta integrator subject to Equations (5) and (7), until equilibrium is reached on an assumed fast time scale. This result is then fed back to $Q_e$, where the contrast-enhanced result is distributed throughout the network.

### 6.3 Static Grouping

Figure 11 represents the response of the CE layer to a static 1-D "image" with structure on multiple scales. In this image there are three point-source inputs at nodes 30, 37, and 42. Initially ($t = 1$), the network responds with maxima at each input, then ($t = 8$) only two maxima survive as the two closest sources interact and merge on the diffusion layer and finally ($t = 19$), all three merge, and the network displays only a single maximum on this coarser scale. As $t \rightarrow \infty$ this network generates a maximum that can be used as a focus of attention near the geometric mean of the input locations. Note that if one's eyes followed the local maxima, concentration would first be on the small, then the larger-scale interactions.

### 6.4 Dynamic Grouping

The following experiments are performed with the time-varying input and use the dynamic DEB parameters listed in Table 1.

#### 6.4.1 Gamma Motion

In this experiment a single light is illuminated, then extinguished. Recall that the percept is of the light expanding upon illumination and contracting when extinguished, with the expansion larger than the contraction. In the simulation the total activity distributed throughout the system is depicted in Figure 12. Activity enters the network through the input endfeet (a), then flows into the central network where activity is diffused into adjacent glial cells (b). This then spreads into the glial enhancement endfeet (c), which is then contrast enhanced (d). In this simulation the activity maximum does not move and thus does not contribute to the motion percept, but the activity edges expand when the dynamic input is first injected into the DEB network, and the activity edges contract when the input is extinguished (see contour plot in Figure 12). It is the motion of these activity edges in the CE layer that yields the motion percept. The relative size of the expansions and contractions agree with the percept.

#### 6.4.2 Long-Range Apparent Motion

In this experiment two lights are illuminated at different locations and times, and when their separation and timing is "correct" the percept is of smooth motion from the first light to the second. In this simulation (Figure 13), both the moving activity maximum and activity edges on the CE layer contribute to the motion percept. The maximum indicates the location of the moving illusory light between the two illumination points. Notice how reenergizing the CE layer upon second input amplifies the activity at the first input. Still the activity maximum moves smoothly between the first and second input locations. The simulation shown here is for stimuli appearing 10 nodes apart, but the same simulation can be run (with no parameter changes) for separations from 2 to 10 nodes apart. Slightly greater distances produce incomplete movement from the first to the second stimuli, while still greater distances produce no movement at all. No movement occurs when the activity

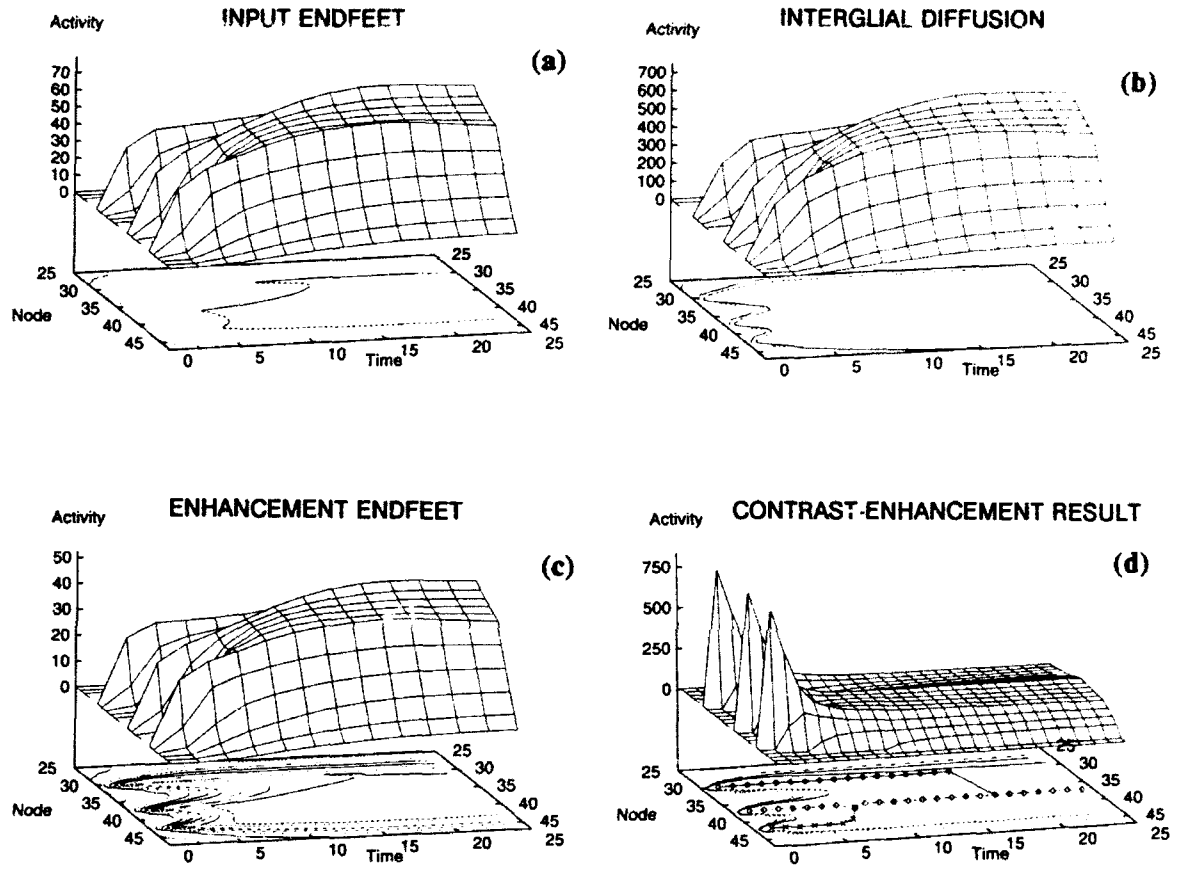


*Figure 11. Multiscale static feature grouping. Activity distributed (a) on glial input endfeet, (b) within interglial layer, (c) on glial enhancement endfeet, and (d) after CE processing. Three inputs at nodes 30, 37, and 42 interact on increasing scale with increasing time. Local maxima trace the grouping of stimuli over time and can serve to focus attention on clusters of features. There is no noise added to this simulation. Contours are only plotted up to 200 units of activity.*

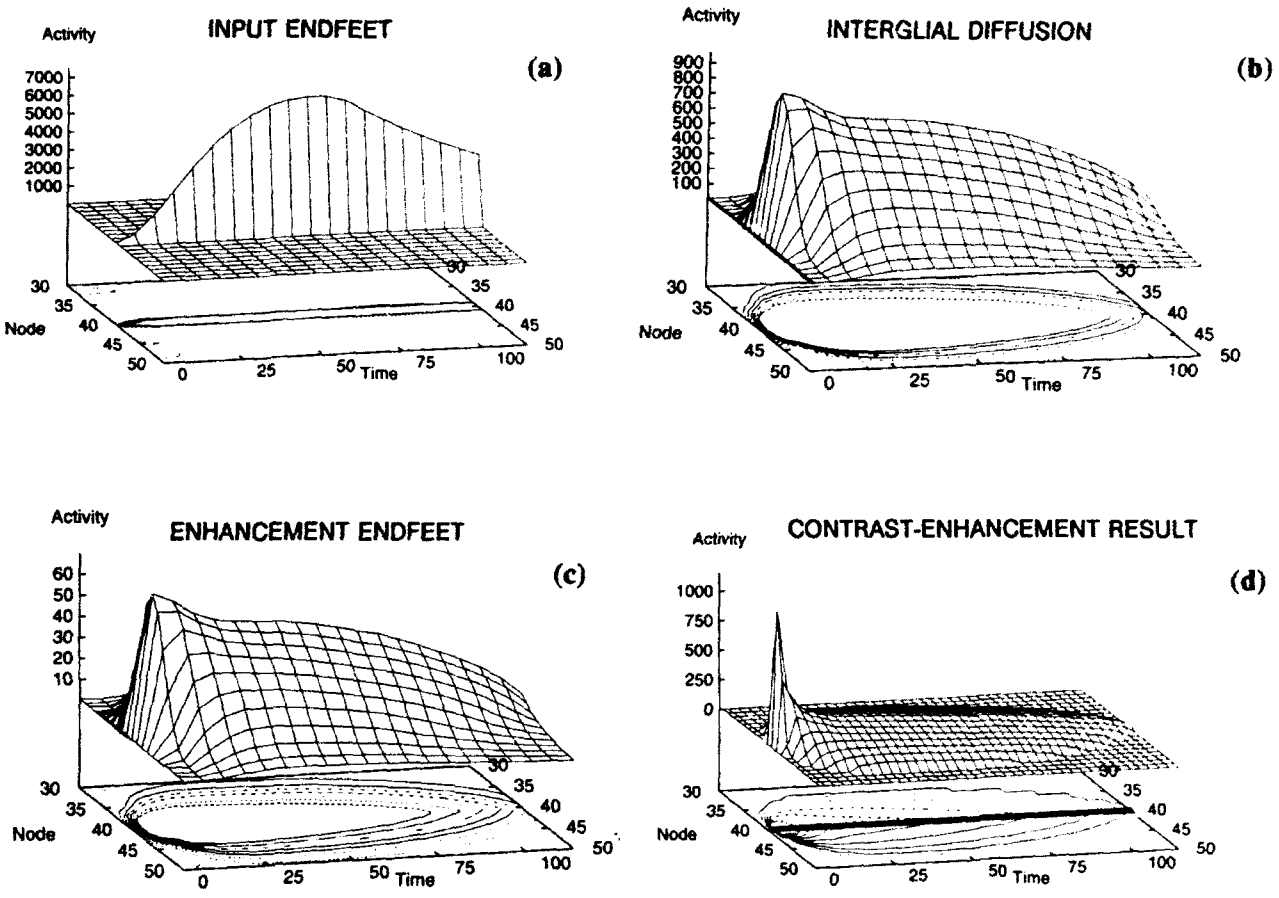


*Figure 12. Gamma motion. Activity distributed (a) on glial input endfeet, (b) within interglial layer, (c) on glial enhancement endfeet, and (d) after CE processing. The CE activity first expands with stimulus onset and then contracts with stimulus offset. There is no motion of the activity maximum, so this does not contribute to the motion percept. The evolution of activity with time across a 1-D chain of nodes is shown, as well as equiactivity contours plotted beneath the activity surfaces. The presence of simulated noise is visible on the input endfeet graph as weak (dashed) contours, but they are suppressed at the output of the CE network.*

"mountains" leak from the system before spreading far enough to interact. This is analogous to the noninteracting percept in Figure 3.

Incomplete movement occurs when the initial local activity maximum moves only part of the way to the distant second maximum, after which the resulting grouping is similar to what is found in the "static" grouping examples.

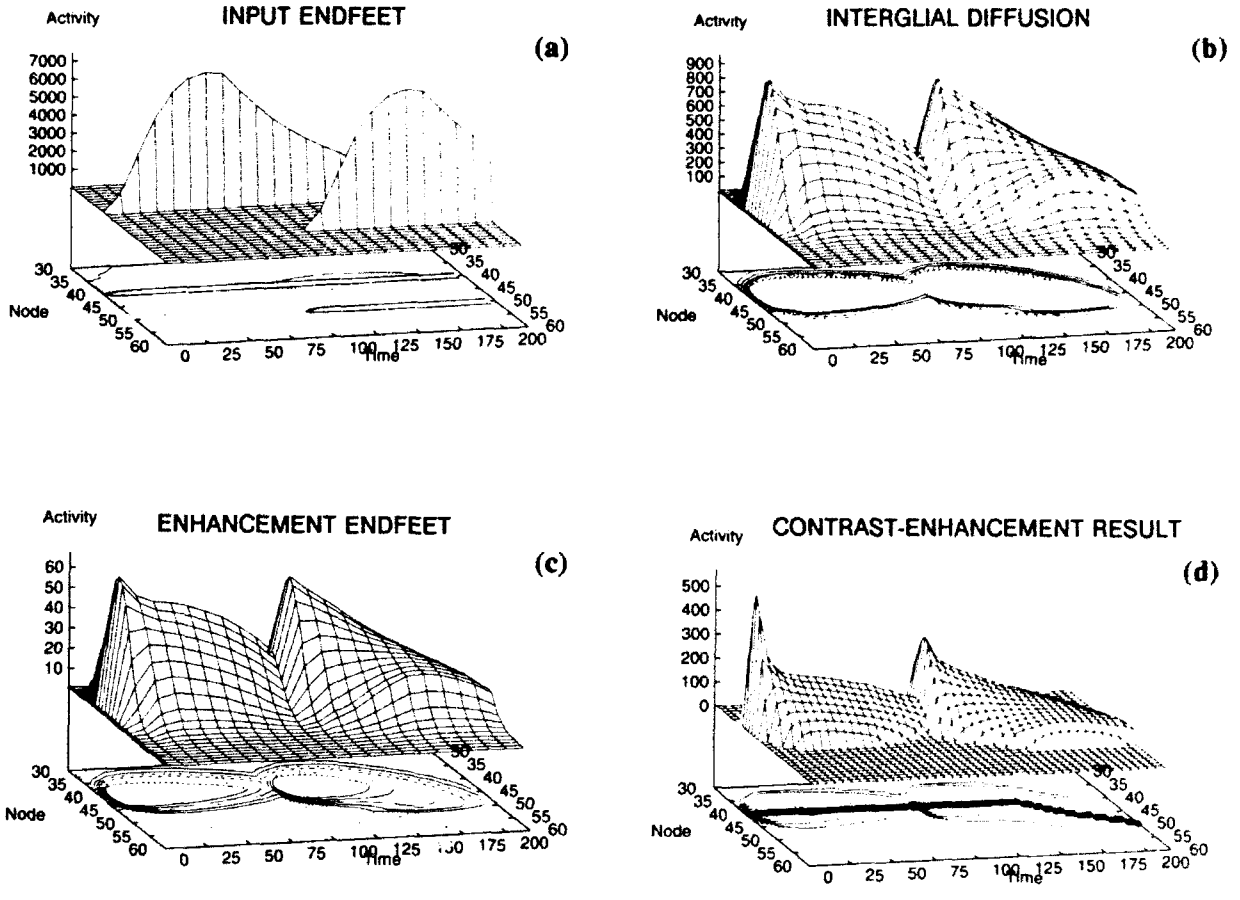


*Figure 13. Long-range apparent motion between two inputs separated in space and time. Activity distributed (a) on glial input endfeet, (b) within interglial layer, (c) on glial enhancement endfeet, and (d) after CE processing. Both the CE activity maximum and the activity edges support the percept of motion from node 40 to node 50, while the activity maximum provides the location of the object in motion. Input is to node 40 at time 0 and node 50 at time 100.*

### 6.4.3 Motion Regimes

By systematically varying the spatial separation and SOA of the inputs, the spatial and temporal intervals can be determined over which various kinds of grouping phenomena can occur. The top curve of Figure 14 delineates the boundary between regimes of smooth motion between first and second stimuli and successive simulated perception of the two stimuli. The bottom curve indicates the boundary between regimes of simulated smooth motion and simulated perception of simultaneity. Just under that line the motion is jumpy at times, while for larger SOAs the motion is smooth. Although there is no jumpy motion at the succession/smooth motion boundary in this simulation, smaller amounts of input activity can induce such jumps.

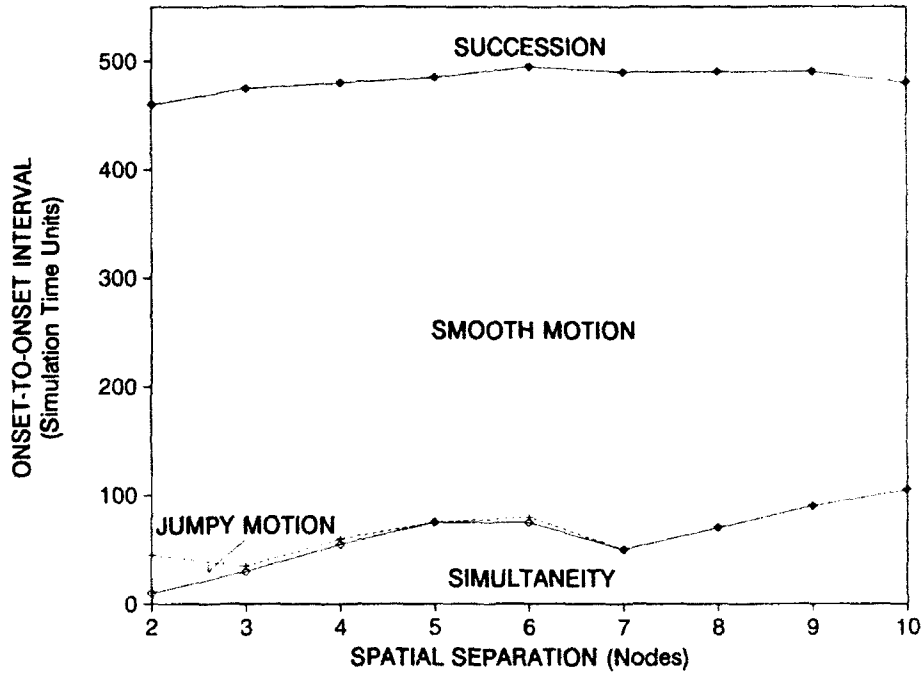


*Figure 14. DEB "cortical" apparent motion regimes.*

To compare the motion regimes shown in Figure 14 with human visual apparent motion perception, each axis should be remapped. Recall that the simulated spatial units (nodes) represent distance in visual cortex, which is a nonlinearly transformed representation of the visual world. Schwartz [48,61] suggested that the mapping from visual angle $z$ (measured from the fovea) to "cortical" displacement $w$ is closely approximated by $w = \log(z + a)$, where $a \approx 1$ deg for rhesus

and squirrel monkeys. Figure 15 uses the inverse map with the end points fit to the approximate spatial separations used by Kolers [4]. With this mapping, a separation of 2 "cortical" nodes represents 0.5 deg of visual angle, while a separation of 10 "cortical" nodes represents about 4 deg.

The time units of the temporal axis of Figure 15 compares closely with milliseconds of real time, in part because the DEB parameters were chosen so that the lower curve would occur in approximately the correct place. The upper curve starts slightly higher than the actual data (which begins around 390 ms), but because the difference is small it was not remapped.

Figure 15 shows the result of this mapping on the motion regimes obtained with the DEB model. This figure agrees qualitatively with human perceptual data as shown in Figure 4. Increased separation between stimuli produces smooth motion over a decreased range of SOA because both the upper and lower bounds converge. The lower curve "predicts" Korte's law. There are, however, quantitative differences: both upper and lower bounds found in the simulations do not converge as rapidly as the real data, and both curves obtained from perceptual data are monotonic, while the simulated data contains minor perturbations.

The simulated regimes are also in agreement with the perceptual data: below the bottom curve, both stimuli simultaneously produce local maxima; immediately above the bottom curve, individual nodes are skipped along the path from first to second node (jumpy motion); above this curve smooth motion is produced; above the upper curve stimuli are not grouped by the network, and the output is two separate maxima.

### 6.4.4 Simulation Velocity

Data presented in Section 6.4.3 can be used to generate apparent motion velocity curves. Figure 16 illustrates the simulated apparent motion velocity. If these data are remapped to "retinal" coordinates using the same mapping as presented in Section 6.4.3, Figure 17 results.

Figure 17 is qualitatively similar to the psychophysical velocity curves depicted in Figure 5. In the psychophysical curves, the upper boundary to the smooth motion regime has a maximum that occurs at about 2.3 deg, which is in agreement with the simulation. The slope of the bottom boundary for experimental and simulated data is positive, but the simulation result has a slope approximately two-thirds that of the experimentally determined slope. Similarly, the slopes of the upper boundary are both positive, but the simulated slope is about twice the experimental slope, resulting in larger apparent velocities than observed. Finally, the upper curve's local peak near 0.75 deg does not occur experimentally.

### 6.4.5 Reverse Motion

In this experiment the first stimulus is presented for a much longer duration than the second stimulus. Motion is initially perceived toward the second stimulus, then back toward the first. In the corresponding simulation shown in Figure 18, the short time span of the second stimulus is assumed to truncate the input of the second stimulus after only 30 time units, perhaps due to prior

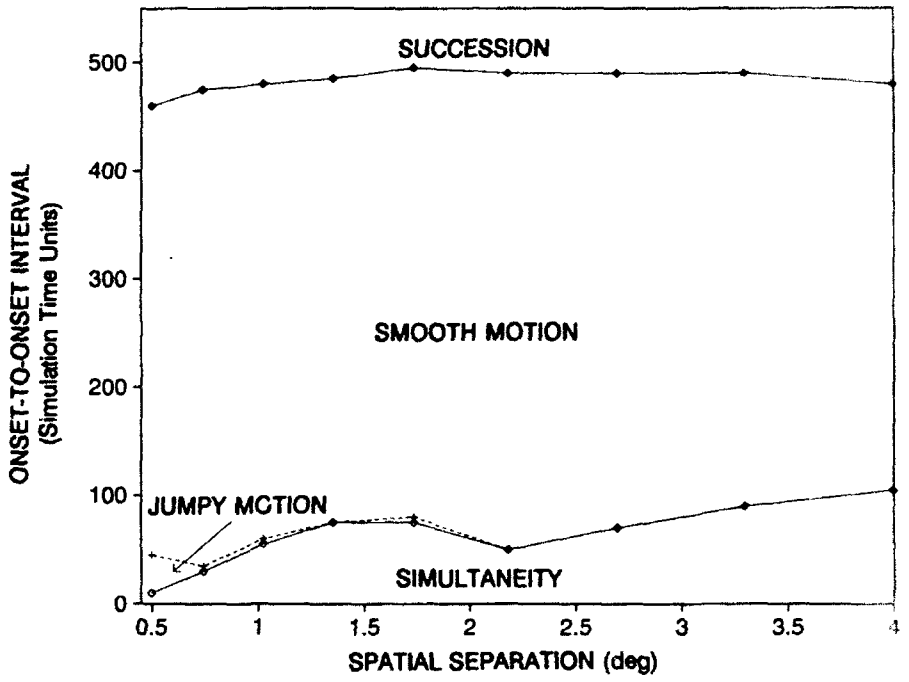


*Figure 15. DEB "retinal" apparent motion regimes. The regimes found from model simulations are in qualitative agreement with the perceptual data. Below the bottom curve, both stimuli simultaneously produce local maxima; immediately above the bottom curve, individual nodes are skipped along the path from first to second node (jumpy motion); above this curve smooth motion is produced; above the upper curve stimuli are not grouped by the network, and the output is two separate maxima. End points of 0.5 and 4 deg have been fit to simulated separations of 2 and 10 nodes, respectively, accounting for a logarithmic mapping from "retinal" to "cortical" distances.*

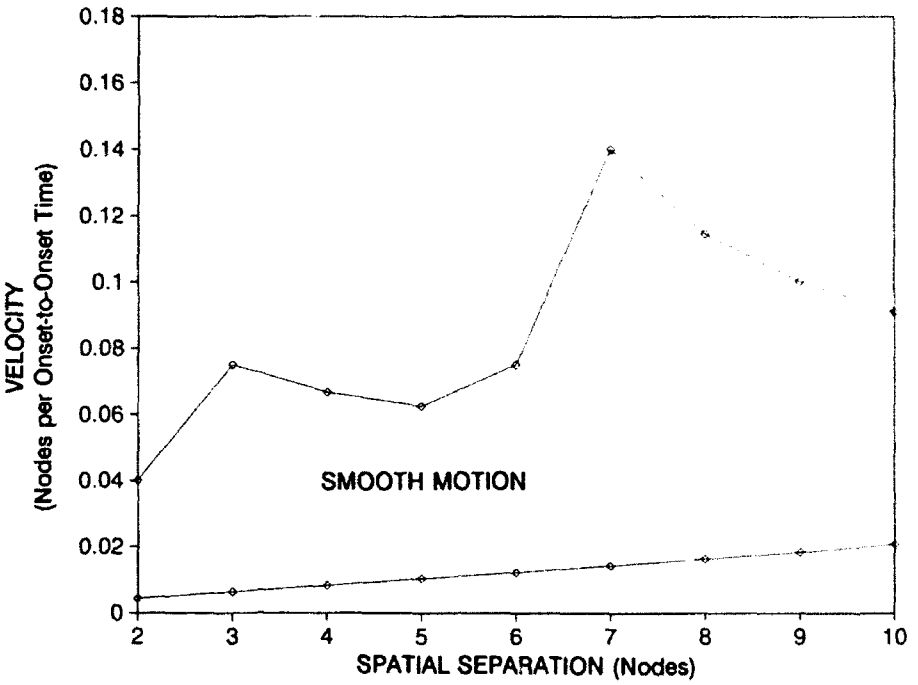


*Figure 16. DEB "cortical" apparent motion velocity versus spatial separation in nodes (i.e., "cortical" distance).*

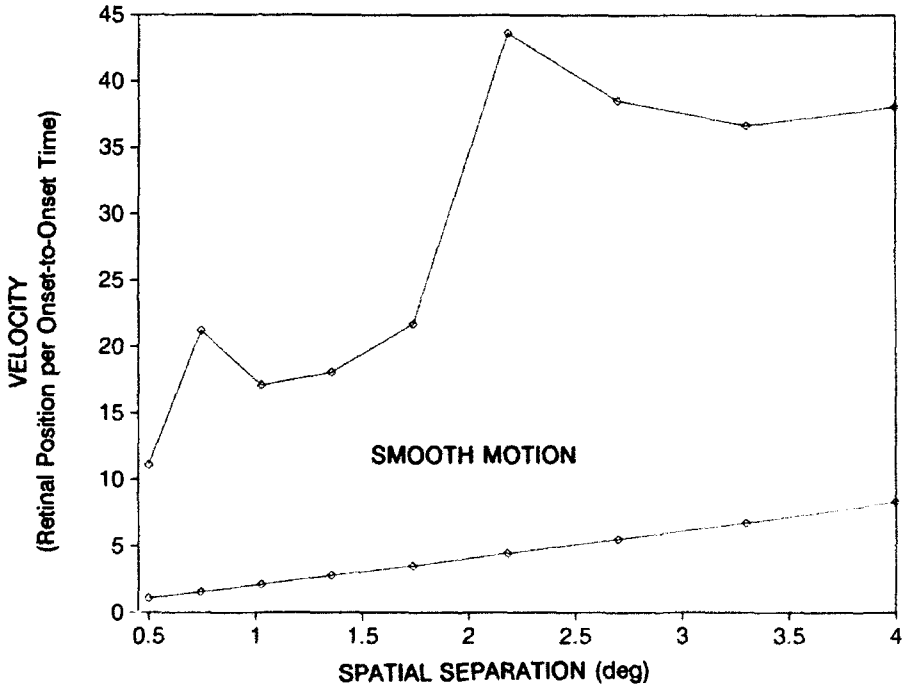


*Figure 17. DEB "retinal" apparent motion velocity versus spatial separation.*

competition between ON and OFF channels. Input is to node 40 at time 0 and node 47 at time 100 (the curves of Figure 18 plot activity from time 150 onward). Smooth motion of the local maximum is found forward from node 40 to 43, before reversing direction and moving back to node 42. In this simulation the weak second input does not generate sufficient activity in the diffusion layer to draw over the hump of activity produced by the stronger first input.

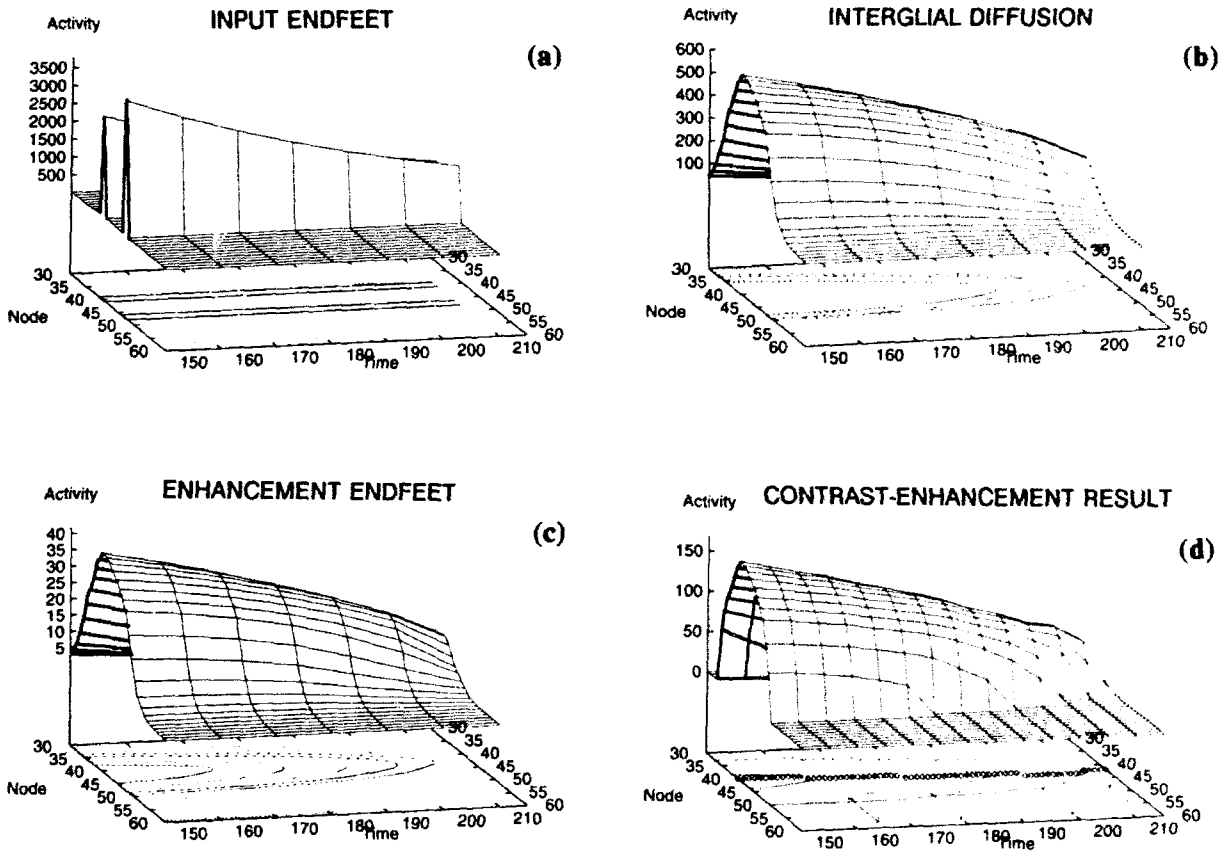


*Figure 18. Simulated reverse motion between two inputs. Activity distributed (a) on glial input endfeet, (b) within interglial layer, (c) on glial enhancement endfeet, and (d) after CE processing. The combination of an extremely long first stimulus and an extremely short second stimulus generates motion of the local maximum partway toward the second input but then begins to move backward toward the initial input.*

### 6.4.6 Equidistant Merge

In this experiment three lights interact; first two lights are flashed, then a third light is flashed equidistant and between the first two. The percept is of the first two lights fusing at the third light's central location. In the simulation depicted in Figure 19, the three lights' effects can be clearly seen in the activation graph of the input endfeet. As the concentrated activity moves from the input endfeet into the interglial communication layer, activity diffuses and begins to interact. This activity moves into the CE endfeet, where the CE layer takes up the activation. Here the merge sensation is reinforced by the motion of activity edges, while the activity maximum moves smoothly from the location of the two initial lights to the central light.[5]

While not illustrated here, the peripheral version of this simulation (in which all lights lie in the periphery) subjects "retinal" input to the "cortical" logarithmic mapping as noted in Section 6.4.3. The simulation of this case results in grouping of the more peripheral light in the first frame with the single light in the second frame, because the logarithmic mapping upsets the equal distances among stimuli in the periphery.

### 6.4.7 Equidistant Split

This experiment reverses the presentation order used in the merge experiment. Again three lights interact; first one light is flashed, then two lights are flashed, equidistant to the first. The percept is of the first light splitting in two with each half moving off to join with either the second or third lights of the second frame. In the simulation shown in Figure 20 (depicted from the time the second frame is introduced) the three lights' effects can be clearly seen in the activation graph of the input endfeet. Here the split sensation is reinforced by the motion of activity edges in the CE layer, which expands significantly with the introduction of the two later stimuli. Unfortunately, three final maxima are produced rather than one maximum splitting and joining the later two, so it is difficult to interpret the sensation of object localization. Eventually, the network groups all inputs together as in the simulation of grouping static inputs.

Input is to node 40 at time 0 and nodes 33 and 47 at time 100. At longer SOAs the network produces significantly more stable local maxima at the outer inputs.

### 6.4.8 Equidistant Peripheral Split/Merge

In this experiment the equidistant split/merge experiments described previously are placed to the side of the fixation point. Perceived motion is to/from the more peripheral light. In this simulation of the split case, displayed are the "cortical" grouping phenomena for which the central,

---

[5]If the third input is removed, static grouping of two inputs is obtained. Running this simulation in the dynamic pathway produces a broader, weaker maximum than that obtained with a two-input merge experiment in the static pathway. Recall the different parameters listed in Table 1.

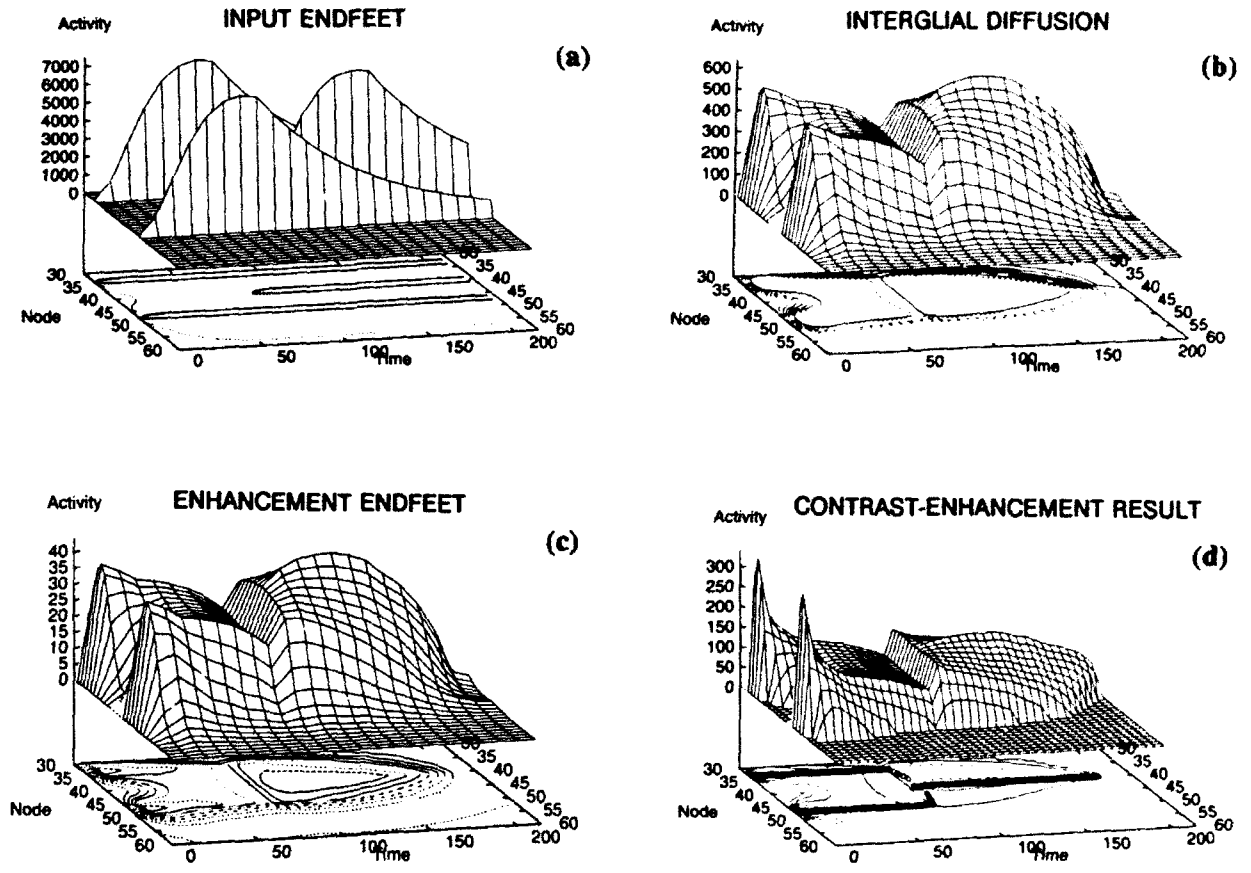


*Figure 19. Equidistant merge. Activity distributed (a) on glial input endfeet, (b) distributed within interglial layer, (c) on glial enhancement endfeet, and (d) after CE processing. Both the CE activity maxima and the activity edges support the motion percept of the two initial stimuli fusing with the central stimulus; the activity maximum provides the location of the stimuli in motion. Input is to nodes 33 and 47 at time 0 and to node 40 at time 80 with no simulated endogenous noise.*

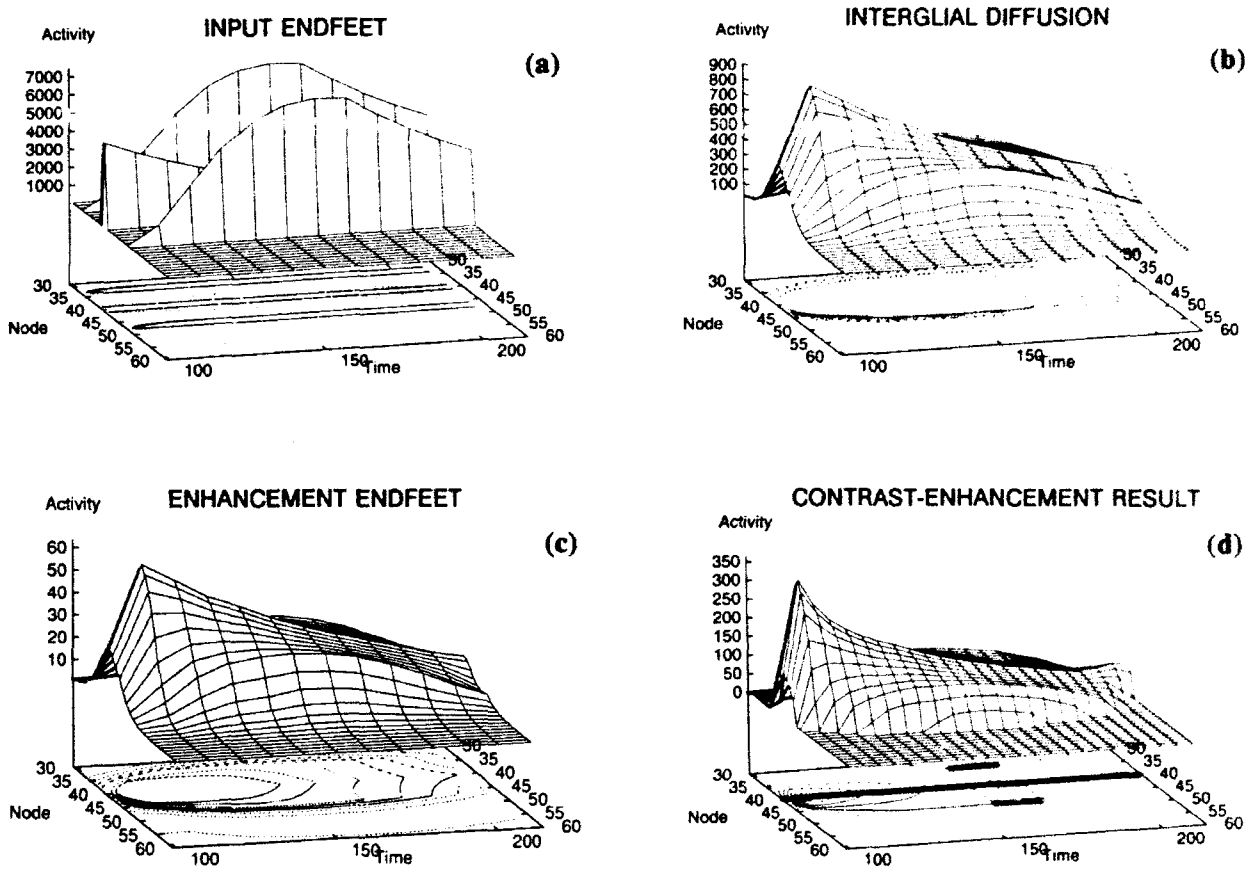


*Figure 20. Equidistant split. Activity distributed (a) on glial input endfeet, (b) within interglial layer, (c) on glial enhancement endfeet, and (d) after CE processing. Both the CE activity maxima and the activity edges support the percept of the central stimulus fissioning and moving off toward the two introduced stimuli. Three activity maxima result (instead of two, as is perceived), corresponding to the location of the stimuli in motion. Eventually, the three maxima are replaced by one central maximum, resulting from the separate activity "mountains" merging.*

first stimulus is 9.6 deg in the periphery (as defined by the mapping used in Section 6.4.3, and the two later stimuli are at 1.7 and 17.4 deg, respectively, corresponding to the later pair of "cortical" stimuli being 9 and 10 nodes from the first stimulus. Figure 21 clearly shows that the initial input moves to the input that is farther from central vision. At a later time, a weak, secondary maximum occurs that can be used to indicate the position of the more central object.

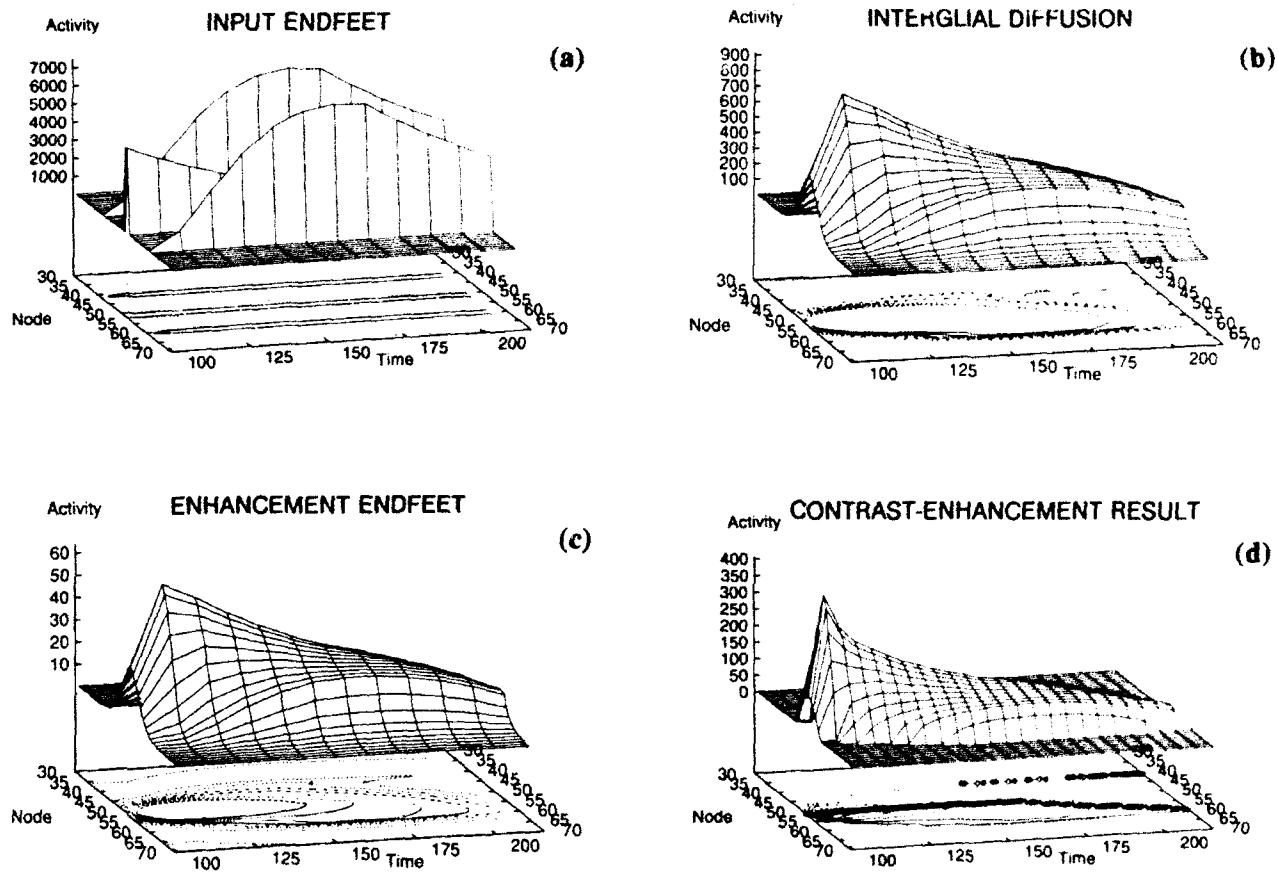


*Figure 21. Equidistant peripheral split. Activity distributed on (a) glial input endfeet, (b) within interglial layer, (c) on glial enhancement endfeet, and (d) after CE processing. Both the CE activity maxima and the activity edges support the percept of motion coming toward the viewer. A very weak maxima at the location of the more central second stimulus periodically occurs but does not move.*

### 6.4.9 Ternus Group Motion

In this experiment three lights move as a group (cf. Figure 8). In the simulation the three lights first coalesce and form one maximum that represents the group, then both activity maxima and activity edges support the sensation of motion to the new group's centroid. There is a sufficiently large ISI in Figure 22 that the entire group of the first frame reasserts itself at the onset of the second frame.

The central maximum represents the coalesced members of the group, which jumps to the side as one member disappears and a new member arrives in the second frame. Although the group as a whole is perceived to move, the individual elements do not disappear from view. This perception is considered the motion analog of the seeing/recognizing dichotomy first discussed by Grossberg and Mingolla [25]. They noted that it was possible to "recognize" static shapes that one did not "see" (e.g., glass patterns, composed of dot pairs, cause one to "recognize" circular patterns that are not "seen"). In this example the dot triplet, moving as a unit, causes one to recognize a rigid object that is not seen.

### 6.4.10 Ternus Element Motion

Element motion in the Ternus configuration is observed at short ISIs and is often considered to be short-range motion. Nevertheless, element motion can also be achieved with the DEB network, if assumptions similar to those made by Grossberg and Rudd [40] are adopted. Those assumptions are that (1) the product of transient and sustained responses to stimuli are input to the DEB network, (2) at small ISI the transient cells are not excited by the central two inputs. (These assumptions agree with psychophysical results reported by Pantle and Petersik [45] but disagree with physiological results reported by Schiller and Logothetis [60].) The response resulting from the product of the sustained and transient cells is then input to the DEB model (see Figure 23).

The Ternus stimulus is depicted in Figure 8, and a cartoon of the CE layer response is superimposed on the DEB input in Figure 24. When the three lights come on in Frame 1, three separate responses (depicted as dots) are generated at the output of DEB's CE layer. Each input supplies activity to the network, which spreads activity over an increasingly wide area. Eventually, activity distributions superimpose, and a single maximum is located at the central input. The first frame ends, a short ISI occurs, and the second frame begins. According to assumption (2), transients at the central two input locations are not excited; however, the OFF response of the bottom input in Figure 24 is excited, and because it more rapidly delivers charge than the ON response (cf. Figure 23), motion is from the central maxima representing the group to the newly excited OFF response of the bottom maxima.

As the slowly rising ON response of the top input continues to be a source of new activity, the local maximum is drawn from the bottom to the top, i.e., between the extreme ends of the display. Simulations have produced motion from the central, group maximum to the top input, but motion to the bottom input of Figures 8 and 24 has proved difficult to generate in the parameter regime used for the other experiments. These simulations only work when long ISIs are used, but

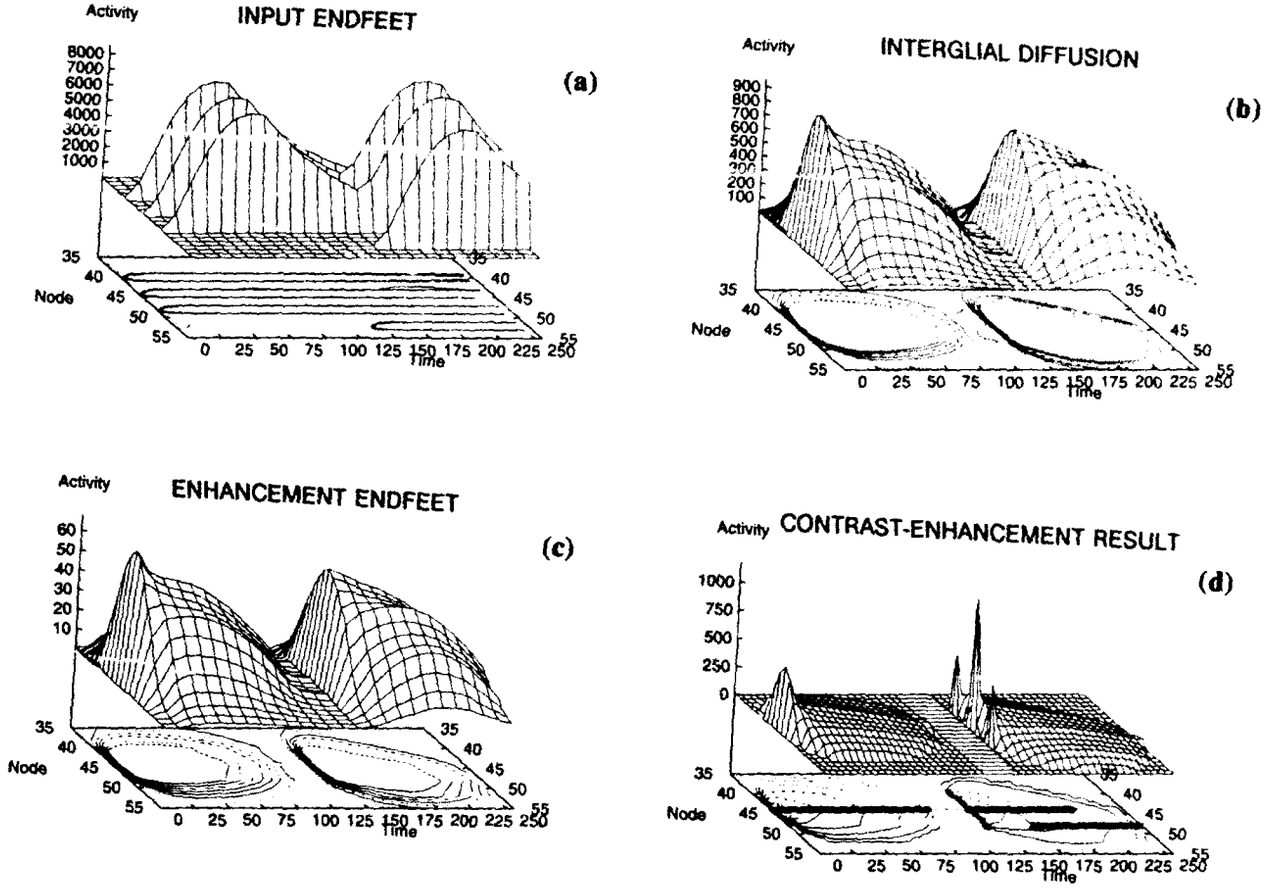


*Figure 22. Simulated Ternus group motion. Activity distributed (a) on glial input endfeet, (b) within interglial layer, (c) on glial enhancement endfeet, and (d) after CE processing. Three local maxima induced by three separate inputs merge together and form a group. This group persists until the second triplet (presented in the second frame) forms a group the activity of which exceeds that of the first group. Motion is from the first to the second group.*

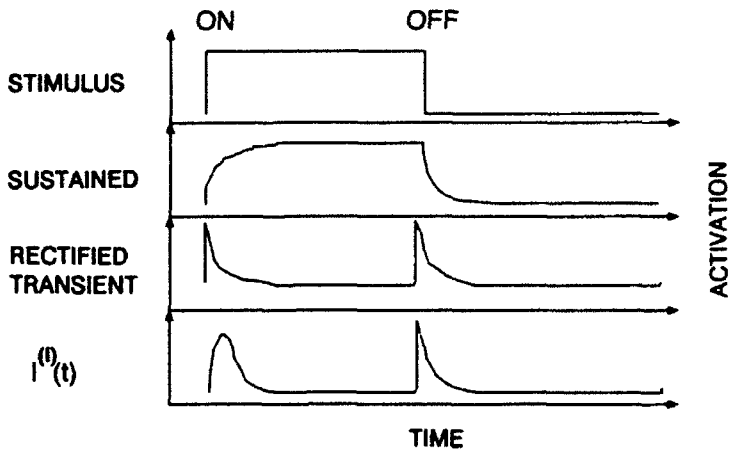


*Figure 23. Sustained and transient cell responses. A point stimulus is turned on and off. The sustained cell activates slowly and remains excited while the stimulus is on. The transient rapidly fires at stimulus onset and offset. The product of the sustained and transient cells is used as input to the DEB network. Notice that the input for the ON response delivers less initial charge than the input for the OFF response (after Grossberg and Rudd [40]).*

this is incompatible with the psychophysical result that short ISIs produce this kind of apparent motion and assumption (2), stated earlier. Thus the DEB model seems unable to adequately reproduce Ternus element motion, which may, in fact, be correct, inasmuch as element motion can be supported directly by a short-range motion process.

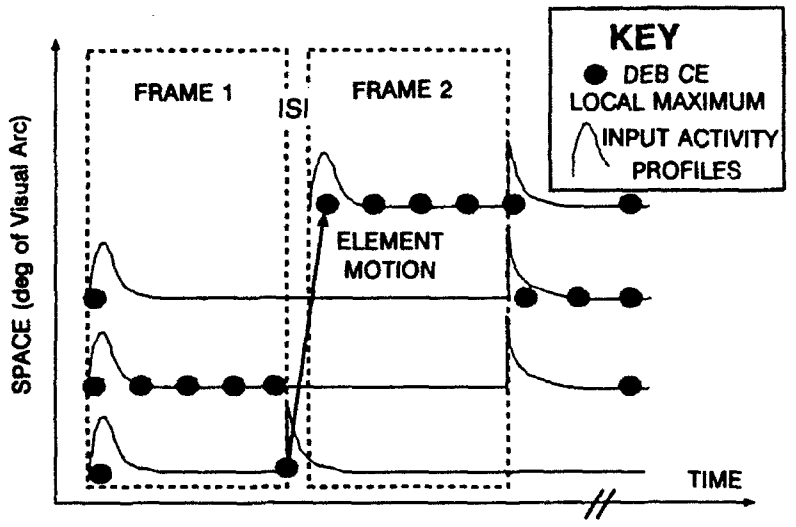


*Figure 24. Ternus element motion explanation. The two central spots do not excite the transient cells between stimulus presentations so no new activity enters the system from these inputs when Frame 1 ends or Frame 2 begins. Motion is from the OFF response of the bottom input to the ON response of the top input.*

# 7. DISCUSSION

A dynamic model for a bidirectionally coupled, two-layered network that spatiotemporally groups its inputs on multiple scales as a function of time has been explored, and the role of other motion models has been considered. The architecture of this diffusion-enhancement network is quite distinct from any large-scale versions of traditional short-range motion detection systems. It thus provides a possible model for a separate long-range motion system, as was originally proposed.

In the DEB model, long-range apparent motion phenomena are generated between spatially and temporally distinct inputs in the presence of endogenous noise. Furthermore, the DEB can group static inputs on multiple scales over time. The output of this grouping process is then input to a short-range motion model. As noted in Section 2.1, the distinction between short- and long-range systems has recently been blurred. This new model, which uses a short-range model as the postprocessor to a long-range model, suggests that the long-range motion system should at times, behave like the short-range system. The relative effects of motion aftereffects is one phenomenon that can be explained with this model. Originally these aftereffects were only associated with short-range stimuli, but more recently, a weak effect has been found by examining bias in long-range motion stimuli [21]. In this experiment a motion aftereffect has been measured in the space between the inducing long-range stimuli, which is exactly what the DEB model predicts: the short-range system that follows the long-range spatiotemporal grouping should adapt to moving activity edges and maxima in exactly the same manner as it would when receiving input from the luminance domain. The difference in the strength of the effect could be due to the relative densities of the input features. In the long-range apparent motion version, only one or a few inputs are smoothly moving in a coherent direction; while in the short-range, luminance-based examples, hundreds of points are moving coherently.

The fact that the DEB model feeds a short-range motion module and that long-range motion can adapt this short-range process, suggests that both DEB output and motion sensing neurons may feed the very same short-range motion module. This scenario is consistent with the known anatomy of early visual processing, discussed in Section 4.1, whereby multiple paths feed the short-range motion area MT (visual area V5), i.e., V1 $\rightarrow$ V5 versus V1 $\rightarrow$ V2 $\rightarrow$ V5 versus V1 $\rightarrow$ V2/V3 $\rightarrow$ V5.

Unfortunately, this model has a few shortcomings that must be addressed. First, we are unable to determine if diffusive interactions in vivo occur at a rate compatible with apparent motion data. In vitro experiments with glial cells examine the time to pass through a cell, but they do not indicate the time for influence via electric fields to be felt at a distant point along the cell. New experiments or calculations should be done to determine if this is feasible. Second, multiple presentations of long-range apparent motion stimuli generate the percept of oscillatory motion between the two end points. If the DEB network parameters are set as stated earlier, and the stimuli are presented multiple times, the network will begin to saturate. In the model presented here, the passive decay term $\frac{G_g}{C_g} Q_g^{(i)}$ is responsible for removing activity from the diffusion network. In the presented

simulations, $\frac{G_g}{C_g}$ is not large enough to drain the system in one presentation cycle. An active mechanism that supplies an opponent inhibitory input may need to be added. It is known that two pathways, ON and OFF, exist at least as far as lateral geniculate nuclei [62]; perhaps inhibition between these two pathways could improve the temporal response of the network. Such interacting networks have been successfully used by Gaudiano [63,64] to model spatiotemporal processing in $X$ and $Y$ retinal ganglion cells.

In conclusion, this model provides a new dynamical framework on which to interpret long-range apparent motion phenomena in the visual, auditory, and somatosensory systems.

## REFERENCES


1. R.K. Cunningham and A.M. Waxman, Parametric study of diffusion-enhancement networks for spatiotemporal grouping in real-time artificial vision. Annual Technical Summary ESD-TR-91-108, Lexington, Mass.: MIT Lincoln Laboratory (1991).

2. R.K. Cunningham and A.M. Waxman, Parametric study of diffusion-enhancement networks for spatiotemporal grouping in real-time artificial vision. Annual Technical Summary ESD-TR-92-121, Lexington, Mass.: MIT Lincoln Laboratory (1992).

3. S. Exner, Ueber das sehen von bewegungen und die theorie des zusammengesetzen auges, *Sitzungsberichte Akademie Wissenscharf Wien*, 156–190 (1875).

4. P.A. Kolers, *Aspects of Motion Perception*, New York: Pergamon Press (1972).

5. M. Wertheimer, Experimentelle studien über das sehen von bewegung, *Zeitschrift für Psych.*, 161–265 (1912), translated in part in *Classics in Psychology*, T. Shipley, ed., New York: Philosophical Library (1961).

6. A.S. Bregman, *Auditory Scene Analysis*, Cambridge, Mass.: MIT Press (1990).

7. G.C. Zapparoli and L.L. Reatto, The apparent movement between visual and acoustic stimulus and the problem of intermodal relationals, *Acta Psychologica*, 256–267 (1969).

8. F.A. Geldard and C.E. Sherrick, Space, time and touch, *Sci. Am.* (April 1986).

9. M. Seibert and A.M. Waxman, Spreading activation layers, visual saccades, and invariant representations for neural pattern recognition systems, *Neural Networks*, 9–27 (1989).

10. A.M. Waxman, M. Seibert, R.K. Cunningham, and J. Wu, The neural-analog diffusion-enhancement layer and early visual processing, *Proc. SPIE Conf. Vis. Commun. Image Process. '88*, 1093–1102 (1988).

11. A.M. Waxman, M. Seibert, R.K. Cunningham, and J. Wu, The neural-analog diffusion-enhancemer layer and spatiotemporal grouping in early vision, *Advances in Neural Information Processing Systems*, D.S. Touretzky, ed., San Mateo, Calif.: Morgan Kaufman (1988), pp. 289–296.

12. A.M. Waxman, J. Wu, and M. Seibert, Computing visual motion in the short and the long: From receptive fields to neural networks, *Proc. IEEE Workshop Vis. Motion*, 156–164 (1989).

13. S. Grossberg, Contour enhancement, short-term memory, and constancies in reverberating neural networks, *Stud. Appl. Math.*, 217–257 (1973).

14. O. Braddick, A short-range process in apparent motion, *Vis. Res.*, 839–847 (1974).

15. S.M. Anstis, The perception of apparent movement, *Philis. Trans. R. Soc. Lond.*, **B**, 153–168 (1980).

## REFERENCES
## (Continued)


16. O. Braddick, Low-level and high-level processes in apparent motion, *Philis. Trans. R. Soc. Lond.*, **B**, 137–151 (1980).

17. S.M. Anstis and B.P. Moulden, Aftereffect of seen movement: Evidence for peripheral and central components, *Q. J. Exp. Psych.*, 222–229 (1970).

18. V.S. Ramachandran and R. Gregor, Does colour provide an input to human motion perception?, *Nature*, 55–56 (1978).

19. J.J. Change and B. Julesz, Displacement limits for spatial frequency filtered random-dot cinematograms in apparent motion, *Vis. Res.*, 1379–1385 (1983).

20. P. Cavanagh, J. Boeglin, and O.E. Favreau, Perception of motion in equiluminance kinematograms, *Perception*, 1595–1601 (1985).

21. M. von Grünau, A motion after effect for long-range stroboscopic apparent motion, *Percept. Psychophys.*, 31–38 (1986).

22. P. Cavanagh and G. Mather, Motion: The long and short of it, *Spatial Vis.*, 103–129 (1989).

23. J.T. Petersik and A. Pantle, Factors controlling the competing sensations produced by a bistable stroboscopic motion display, *Vis. Res.*, 143–154 (1979).

24. C. Chubb and G. Sperling, Drift-balanced random stimuli: A general basis for studying non-Fourier motion perception, *J. Opt. Soc. Am.* **A**, 1986–2006 (November 1989).

25. S. Grossberg and E. Mingolla, Neural dynamics of form perception: Boundary completion, illusory figures, and neon color spreading, *Psych. Rev.*, 173–211 (1985).

26. S. Grossberg and E. Mingolla, Neural dynamics of perceptual grouping: Textures, boundaries, and emergent segmentations, *Percept. Psychophys.*, 141–171 (1985).

27. R.K. Cunningham and A.M. Waxman, Astroglial-neural networks, diffusion-enhancement bilayers, and spatio-temporal grouping dynamics, *Proc. Int. Joint Conf. Neural Net.*, 937 (July 1991).

28. R.K. Cunningham and A.M. Waxman, Astroglial-neural networks, diffusion-enhancement bilayers, and spatio-temporal grouping dynamics, *Proc. SPIE Conf. Sensor Fusion IV: Control Paradigms and Data Structures*, 411–422 (November 1991).

29. E.H. Adelson and J.R. Bergen, Spatiotemporal energy models for the perception of motion, *J. Opt. Soc. Am.* **A**, 284–299 (1985).

30. P. Anandan, A computational framework and an algorithm for the measurement of visual motion, *Int. J. Comput. Vis.*, 283–310 (1989).

## REFERENCES
## (Continued)


31. H.B. Barlow and W.R. Levick, The mechanism of directionally selective units in the rabbit retina, *J. Physiol.*, 477–504 (1965).

32. W. Reichardt, Autocorrelation, a principle for the evaluation of sensory information by central nervous system, *Sensory Communication*, W.A. Rosenblith, ed., New York: Wiley Press (1961).

33. J.P.H. van Santen and G. Sperling, Elaborated Reichardt detectors, *J. Opt. Soc. Am.* **A2**, 300–321 (1985).

34. K.P. Horn and B.G. Schunch, Determining optical flow, *Artif. Intell.*, 185–203 (1981).

35. D. Marr and S. Ullman, Directional selectivity and its use in early visual processing, *Proc. R. Soc. Lond.* **B**, 151–280 (1981).

36. S. Ullman, Analysis of visual motion by biological and computer systems, *Comput. Mag.*, 57–69 (August 1981).

37. D. Fay and A.M. Waxman, Neurodynamics of real-time image velocity extraction, *Neural Networks for Vision and Image Processing*, G.A. Carpenter and S. Grossberg, eds., Cambridge, Mass.,: MIT Press (1992), pp. 221-246.

38. A.M. Waxman, J. Wu, and F. Bergholm, Convected activation profiles and the measurement of visual motion, *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, 717–723 (1988).

39. S. Ullman, *The Interpretation of Visual Motion*, Cambridge, Mass.: MIT Press (1982).

40. S. Grossberg and M.E. Rudd, A neural architecture for visual motion perception: Group and element apparent motion, *Neural Net.*, 421–450 (1989).

41. J.L. Marroquin, Human Visual Perception of Structure, Master's thesis, Cambridge, Mass.: Massachusetts Institute of Technology (1976).

42. S.H. Bartley, *Vision: A Study of Its Basis*, New York: D. Van Nostrand Company, Inc., (1941) Chapter VII.

43. W. Neuhaus, Experimentelle untersuchung der scheinbewegung, *Archiv für die gesamt Psychologie*, 315–458 (1930).

44. J. Ternus, The problem of phenomenal identity, *A Source Book of Gestalt Psychology*, W.D. Ellis, ed., Routledge and Kegan Paul (1938).

45. A. Pantle and J.T. Petersik, Effects of spatial parameters on the perceptual organization of a bistable motion display, *Percept. Psychophys.*, 307–312 (1980).

46. A. Pantle and L. Picciano, A multistable movement display: Evidence for two separate motion systems in human vision, *Science*, 500–502 (1976).

## REFERENCES
## (Continued)


47. D.H. Hubel, *Eye, Brain and Vision*, New York: W.H. Freeman & Co., Scientific American Library (1988).

48. E.L. Schwartz, Spatial mapping in the primate sensory projection: Analytic structure and relevance to perception, *Biol. Cybernetics*, 181–194 (1977).

49. M. Livingstone and D. Hubel, Segregation of form, color, movement, and depth: anatomy, physiology, and perception, *Science*, 740–749 (1988).

50. C. Mason and E.R. Kandel, Central visual pathways, *Principles of Neural Science*, 3rd Ed., E.R. Kandel, J.H. Schwartz, and T.M. Jessell, eds., New York: Elsevier (1991), pp. 420–439.

51. S. Zeki and S. Shipp, The functional logic of cortical connections, *Nature*, 311–317 (September 1988).

52. H.K. Kimelberg and M.D. Norenberg, Astrocytes, *Sci. Am.*, 91–96 (April 1989).

53. H. Kettenman and B.E. Ransom, Electrical coupling between astrocytes and between oligodendrocytes studied in mammalian cell cultures, *Glia*, 64–73 (1988).

54. J.A. Conners, L.S. Benardo, and D.A. Prince, Carbon dioxide sensitivity of dye coupling among glia and neurons of the neocortex, *J. Neurosci.*, 1324–1330 (1984).

55. S.W. Kuffler, J.G. Nicholls, and R.A. Martin, *From Neuron to Brain*, 2nd Ed., Sunderland, Mass.: Sinauer Associates, Inc. (1985).

56. L.L. Odette and E.A. Newman, Model of potassium dynamics in the central nervous system, *Glia*, 198–210 (1988).

57. A.H. Cornell-Bell, S.M. Finkbeiner, M.S. Cooper, and S.J. Smith, Glutamate induces calcium waves in cultured astrocytes: Long-range glial signaling, *Science*, 470–473 (January 1990).

58. L. Hertz, Possible role of neuroglia: A potassium-mediated neuronal-neuroglial-neuronal impulse transmission system, *Nature*, 1091–1094 (1965).

59. C. Enroth-Cugell and J.G. Robson, The contrast sensitivity of retinal ganglion cells of the cat, *J. Physiol.*, 527–552 (1966).

60. P.H. Schiller and N.K. Logothetis, The color-opponent and broad-band channels of the primate visual system, *TINS*, 392–398 (1990).

61. E.L. Schwartz, Computational anatomy and funtional architecture of striate cortex: A spatial mapping approach to perceptual coding, *Vis. Res.*, 645–669 (1980).

62. P.H. Schiller, The connections of the retinal on- and off- pathways on the lateral geniculate nucleus of the monkey, *Vis. Res.*, 923–932 (1984).

## REFERENCES
## (Continued)


63. P. Gaudiano, A unified neural network model of spatiotemporal processing in $X$ and $Y$ retinal ganglion cells, I: Analytical results, *Biol. Cybernetics*, 11–21 (1992).

64. P. Gaudiano, A unified neural network model of spatiotemporal processing in $X$ and $Y$ retinal ganglion cells, II: Temporal adaptation and simulation of experimental data, *Biol. Cybernetics*, 23–34 (1992).

# REPORT DOCUMENTATION PAGE

*Form Approved*
*OMB No. 0704-0188*



| 1. AGENCY USE ONLY (*Leave blank*) | 2. REPORT DATE | 3. REPORT TYPE AND DATES COVERED |
|---|---|---|
| | 6 April 1993 | Final Technical Summary |

**4. TITLE AND SUBTITLE**

Parametric Study of Diffusion-Enhancement Networks for Spatiotemporal Grouping in Real-Time Artificial Vision

**5. FUNDING NUMBERS**

C — F19628-90-C-0002
PR — 388

**6. AUTHOR(S)**

Robert K. Cunningham and Allen M. Waxman

**7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)**

Lincoln Laboratory, MIT
P.O. Box 73
Lexington, MA 02173-9108

**8. PERFORMING ORGANIZATION REPORT NUMBER**

**9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)**

Air Force Office of Scientific Research
Bolling AFB
Washington, DC 20402

**10. SPONSORING/MONITORING AGENCY REPORT NUMBER**

ESC-TR-92-207

**11. SUPPLEMENTARY NOTES**

None

**12a. DISTRIBUTION/AVAILABILITY STATEMENT**

Approved for public release; distribution is unlimited.

**12b. DISTRIBUTION CODE**

**13. ABSTRACT** (*Maximum 200 words*)

Spatiotemporal grouping phenomena are examined in the context of static and time-varying imagery. Dynamics that exhibit static feature grouping on multiple scales as a function of time and long-range apparent motion between time-varying inputs are developed for a biologically plausible diffusion-enhancement bilayer network. The architecture consists of a diffusion layer and a contrast-enhancement layer coupled by feedforward and feedback connections; time-varying input is provided by a separate feature extracting layer. The model is cast as an analog circuit that is realizable in VLSI, the parameters of which are selected to satisfy a psychophysical database of the following long-range apparent motion phemomena: gamma motion of a single light, smooth motion between two lights, reverse motion, split and merge among three lights, Ternus motion among multiple lights, and peripheral motion. The relation between motion on a uniform network (i.e., cortex) and inputs to a nonuniform sampling array (i.e., retina) are discussed in the context of a logarithmic scaling of space. A new interpretation of short- and long-range visual motion systems is introduced.

**14. SUBJECT TERMS**

neural networks
astrocyte glial networks
diffusion enhancement
long-range apparent motion
spatiotemporal grouping dynamics
interference suppression

**15. NUMBER OF PAGES**
68

**16. PRICE CODE**

| 17. SECURITY CLASSIFICATION OF REPORT | 18. SECURITY CLASSIFICATION OF THIS PAGE | 19. SECURITY CLASSIFICATION OF ABSTRACT | 20. LIMITATION OF ABSTRACT |
|---|---|---|---|
| Unclassified | Unclassified | Unclassified | Same As Report |